Infant Behavior and Development 61 (2020) 101475

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

6- to 9-Month old infants discriminate vowel durations in variable speech contexts

Simone Falk[a,*], Christine D. Tsang[b]

[a] Department of Linguistics and Translation, International Laboratory for Brain, Music & Sound Research (BRAMS), University of Montreal, C. P. 6128, Succursale Centre-ville, Montréal Québec H3C 3J7, Canada
[b] Department of Psychology, Huron University College at Western, Canada

ARTICLE INFO

Keywords:
Infant
Speech perception
Phonetic discrimination
Temporal relations
Rhythmic regularity

ABSTRACT

Discriminating temporal relationships in speech is crucial for speech and language development. However, temporal variation of vowels is difficult to perceive for young infants when it is determined by surrounding speech sounds. Using a familiarisation-discrimination paradigm, we show that English-learning 6- to 9-month-olds are capable of discriminating non-native acoustic vowel duration differences that systematically vary with subsequent consonantal durations. Furthermore, temporal regularity of stimulus presentation potentially makes the task easier for infants. These findings show that young infants can process fine-grained temporal aspects of speech sounds, a capacity that lays the foundation for building a phonological system of their ambient language(s).

1. Introduction

Learning about temporal aspects of sounds is crucial for developing speech and communication skills in infants. However, infants face a complex task, as the temporal structure of speech sounds heavily varies within different speech contexts. For example, vowel sounds are longer in phrase-final than in phrase-medial positions to mark the end of sentences or to indicate the moment of communicative turns (e.g., Klatt, 1976; Levinson & Torreira, 2015). Longer vowel durations can also indicate that the speaker provides important or emphatic information (De Jong, 2004). In some languages such as Finnish, Japanese, Hindi or Urdu, long and short durations of vowels mark different entries in the lexicon (i.e., "phonemic length", as in Japanese /kado/ "corner" vs. /ka:do/ "card", Ladefoged, 1996; Mugitani et al., 2009).

It takes time for infants to acquire an understanding of these complex temporal relations and their functions in speech. In non-verbal sounds, infants show discrimination of temporal relations from as early as 4 months of age. Four-month-olds can discriminate small temporal changes, and by 6 months show discrimination for event durations in non-verbal audio-visual sequences (e.g., Brannon, Suanda, & Libertus, 2007; Lewkowicz & Marcovitch, 2006). In audio-only tone sequence, 5-month-olds perceive temporal variations leading to different auditory groupings (e.g., Chang & Trehub, 1977).

Speech-related temporal discrimination starts later. English-learning infants discriminate vowel length differences in one-, two- or three-syllable nonsense words between 5 and 11 months (Eilers, Bull, Oller, & Lewis, 1984). Between 7.5 and 9.5 months, infants learning Japanese start to discriminate vowel length differences which mark lexical entries in their native language, while they can discriminate spectral vowel differences such as /i/ vs. /a/ from 4 months on (Mugitani et al., 2009; Sato, Sogabe, & Mazuka, 2010; Sato, Kato, & Mazuka, 2012). Discrimination of temporal contrasts as a function of language experience and linguistic functions

* Corresponding author.
E-mail address: Simone.falk@umontreal.ca (S. Falk).

https://doi.org/10.1016/j.infbeh.2020.101475
Received 26 September 2019; Received in revised form 23 July 2020; Accepted 23 July 2020

appears to start even later during the second year of life, around 18 months of age (e.g., Mugitani et al., 2009).

It is a possibility that infants, independent of the language they are learning, start with an acoustically driven approach to temporal discrimination of sounds in speech. This idea comes, for example, with the model of infant sound discrimination (Galle & McMurray, 2014; based on Pisoni & Tash, 1974). Their model allows for parallel discrimination modes: continuous perception at the level of acoustical auditory cues and categorical perception at the level of memory-based functional processing. Infants' speech perception would be continually improving during development while the two channels (acoustic and categorical) co-develop to constantly refine infants' discrimination abilities. Younger infants would first discriminate sounds on a predominantly acoustic basis, and older infants increasingly rely on both channels, using both acoustic and categorical discrimination. As a crucial step along this developmental path, infants need to refine their abilities to parse the appropriate (acoustic versus functional) *context* in which a temporal relation occurs. How infants achieve this during language acquisition is still not fully understood.

Indeed, previous findings suggest that infants in their first year of life are not able to acoustically capture temporal vowel variations conditioned by other surrounding sounds. One example is the English "vowel length effect" (Ko, Soderstrom, & Morgan, 2009). Here, vowel duration depends on whether a following consonant is voiced (like in /g/, "pig" /pIg/) or voiceless (as in /k/, "pick" /pIk/). The /i/-vowel followed by a voiced /g/ is consistently longer than the /i/ followed by the voiceless /k/ (e.g., Hillenbrand, Ingrisano, Smith, & Fledge, 1984). The effect is explained by coarticulation (i.e., the transfer of articulatory properties to surrounding speech sounds), as the vowel and consonant durations work together to keep the overall duration of the vowel-consonant pattern constant (note that English voiced plosives are naturally shorter than voiceless ones; Whalen, 1990). By 8 months of age, infants growing up in English-speaking environments are not sensitive to the vowel length effect and show only partial sensitivity to this cue at 14 month (Eilers et al., 1984; Ko et al., 2009). If different combinations of vowels and consonants (e.g., nasals) are used, English-learning infants fail to discriminate temporal differences acoustically in a word-learning task, even at the age of 18 months (Dietrich, Swingley, & Werker, 2007). Only infants (e.g., dutch-learning) whose ambient language features categorical vowel length variation in these contexts are able to discriminate the words as a function of the temporal differences (Dietrich et al., 2007). This indicates that during the second year of life, categorical, language-specific perception may be more important than acoustic perception.

Temporal vowel variations conditioned by a subtle co-articulatory voice contrast may be particularly difficult to perceive acoustically for young infants who themselves have little articulatory practice. Therefore, the present study aims at determining whether infants, in the second half of the first year, are able to acoustically discriminate temporal aspects of vowels that do not depend on subtle co-articulation or consonant quality. We therefore chose a situation in which the temporal structure of vowels systematically varies with other temporal properties of the surrounding speech sounds. In order to find evidence for an acoustic account of infants' temporal perception, we chose a temporal context which was unfamiliar to them in their maternal language. The context is best described as a relation of "complementary durations" between vowels and consonants which is a (sub)part of the phonological system of some languages, such as Swedish or German (Elert, 1964; Falk, 2011). In the German word for 'rats', /rat@n/ ("Ratten"), the short vowel /a/ is followed by an acoustically longer /t/ consonant than in the word for 'guess' /ra:t@n/ ("raten") in which /a:/ is long and /t/ is acoustically shorter.

Based on the previous literature and age reports therein (Eilers et al., 1984), we hypothesize that infants between 6 and 9 months of age will be acoustically sensitive to these vowel length differences even when the unfamiliar temporal context varies. Alternatively, it is possible that infants of this age are increasingly tuned into what is relevant to the ambient language and may not show discrimination for temporal properties of vowels in contexts that typically do not occur in their ambient language (similar to spectral properties, see Tsuji & Cristia, 2014).

Using a familiarization – discrimination paradigm (see Trainor, Wu, & Tsang, 2004), infants were familiarized with repeated pseudo-words featuring different complementary temporal relations between vowels and consonants. They were then tested on a new series of pseudo-words with either the same or a different temporal relation. As stimuli repetition was an integral part of the test paradigm, we controlled effects of overall temporal regularity of word presentation. Note that higher-order temporal regularity in repeated stimulus presentation (i.e., such as a perceived regular "beat" in the word sequence, similar to a musical rhythm) could act as a means to facilitate auditory discrimination tasks to infants (Otte et al., 2013). In particular, infants at the age of our test group, between 5 and 8 months of age, attend longer to regular structured tone sequences than to irregular ones (Nakata & Mitani, 2005), which in turn, may lead to better processing of sound durations in regular sequences compared to irregular ones, similar to adults (Quené & Port, 2005; Zheng & Pierrehumbert, 2010). In sum, we hypothesize that infants between 6 and 9 months are able to discriminate a new contextual temporal relation between vowels and consonants, independently of higher-order temporal regularity in stimulus presentation.

2. Method

2.1. Participants

Fifty-nine 6- to 9-month-old infants (27 males and 32 females, mean age = 7.16 months, SD = 0.76; range = 6.0–9.1 months) participated in this study. An additional 6 infants were tested in the study and completed all trials but were not included in the final sample for analysis due to: infant fussiness during test (n = 3); infant reported with cold (n = 1); incomplete parental questionnaire (n = 2). Participants were recruited via telephone from the developmental research participant database maintained by the BLINDED. Infants were recruited from monolingual English-speaking households in BLINDED. BLINDED is a mid-sized BLINDED city

Table 1
Acoustic characteristics of the stimuli.

Duration ratio between vowels : consonants	Long-short condition: ~2.5 : 1 Short-long condition: ~1: 2
Duration difference between long and short sounds	Vowels: 203 ms (SD = 42 ms) Consonants: 277 ms (SD = 47 ms)
Articulation rate	2 syllables / s
Mean pitch range	220 − 236 Hz

in BLINDED, in which more than 50 % of the adult population has some post-secondary certificate, diploma, or degree, more than 80 % of households indicate English as their native language and more than 90 % of families indicate that English is the predominant language spoken in the home (see BLINDED). Only full-term, normal birthweight infants with no reported issues during the birth process as reported in records from the hospital at the time of birth were eligible to participate. Parents completed a brief infant health questionnaire at the time of testing, and only infants who were healthy with no reported history of ear infections or history of familial hearing loss were included in the study results.

2.2. Stimuli

The stimuli in this study were 4 trisyllabic pseudo-words composed of the syllables “mi” and “la” which were 1.5 s long on average (*SD* = 54 ms). The first syllable was accented and displayed a notable higher pitch than the other two syllables. Vowel-consonant durations in the pseudo-words varied as a function of two conditions. In the long-short condition (/mi:la:mi/ and /la:-mi:la/), the vowels in the first and second syllables were long (mean = 384 ms, *SD* = 18 ms) followed by a short consonant (mean = 152 ms, *SD* = 17 ms). In the short-long condition (/mil:am:i/ and /lam:il:a/), the vowels of the first and second syllables were short (mean = 181 ms, *SD* = 33 ms) and consonants were long (mean = 349 ms, *SD* = 44 ms). A native German-speaking female speaker recorded the stimuli so that the voice was consistent across the presented pseudo-words. Three instances of each pseudo-word were chosen that were most similar in terms of duration ratio, rate, pitch and vowel quality in order to avoid manipulations of the sound (Table 1).

The three versions of each word were concatenated in a quasi-random order that never repeated sequentially. To control for effects of temporal regularity, we introduced a gradual temporal jitter between repetitions. Two conditions were created, a temporally more Regular and a temporally more Irregular condition (see Fig. 1B). We added the latter condition to serve as a control to ensure that infants’ discrimination was not the result of higher-order temporal structure (such as the presence of a perceived beat) during repeated presentation. The temporally more Regular condition was created by setting the inter-word-intervals (IWI) to 2.3 s on average, varying with a slight jitter (*SD* = 180 ms). In the temporally more Irregular condition, we modified the regular stimuli such that the word onset between non-words was much more unpredictable. Therefore, we added +/−0 ms, +/−300 ms, +/−420 ms, +/−310 ms, +/−150 ms, +/−600 ms to the pause between words in the Regular condition (see Fig. 1B). Overall IWIs were of the same duration on average (2.3 ms), but jitter was doubled by the aforementioned procedure (*SD* = 360 ms). There were 32 infants tested in the Regular condition. The Irregular condition was considered a control condition and we tested an additional 27 infants in this condition.

2.3. Procedure

Infants were tested individually. There were two phases in the Experiment: an Exposure (familiarization) Phase and a Test Phase (see Fig. 1A). During the Exposure Phase, the caregiver and infant were brought into the test room. The caregiver was instructed to sit on a chair between two sound speaker cabinets and facing the desk of the experimenter. The caregiver held the infant on the lap and was fitted with headphones that played music to mask the sound of the stimuli directed to the infant. The experimenter then left the room and started the familiarization stimulus for the infant.

The familiarization stimulus was presented in stereo, emanating from both speakers (the right and the left side) and was played using the iTunes application on the Mac Mini computer in the adjoining test room. The infant listened to the stimulus, which consisted of one set of 4 pseudo-words with a specific duration pattern (e.g., /mi:la:mi/ : long-short) for approximately 2 min. Infants were randomly assigned to one of the four familiarization conditions (/la:mi:la/, /lam:il:a/, /mi:la:mi/, /mil:am:i/). During the Test Phase, the infant was presented with the other pseudo-word of the same condition (i.e., long-short, /la:mi:la/) or the alternate condition (i.e., short-long, /lam:il:a/, see Fig. 1A as an example). After the presentation of the familiarization clip, the experimenter re-entered the test room, sat behind the desk directly across from the infant and began the Test Phase when the infant was facing forward and appeared to be attentive.

During the Test Phase, a standard head-turn preference procedure was used (see Kemler Nelson et al., 1995; Tsang, Falk., & Hessel, 2017). The Test Phase began with the image of Mickey Mouse flashing on a computer screen to one side (right or left side) of the infant. When the infant turned his/her head to look at the screen, the target image stopped flashing and remained on the screen as the experimenter pressed a key from behind the desk that prompted one of the sound stimuli to begin playing from the sound speaker

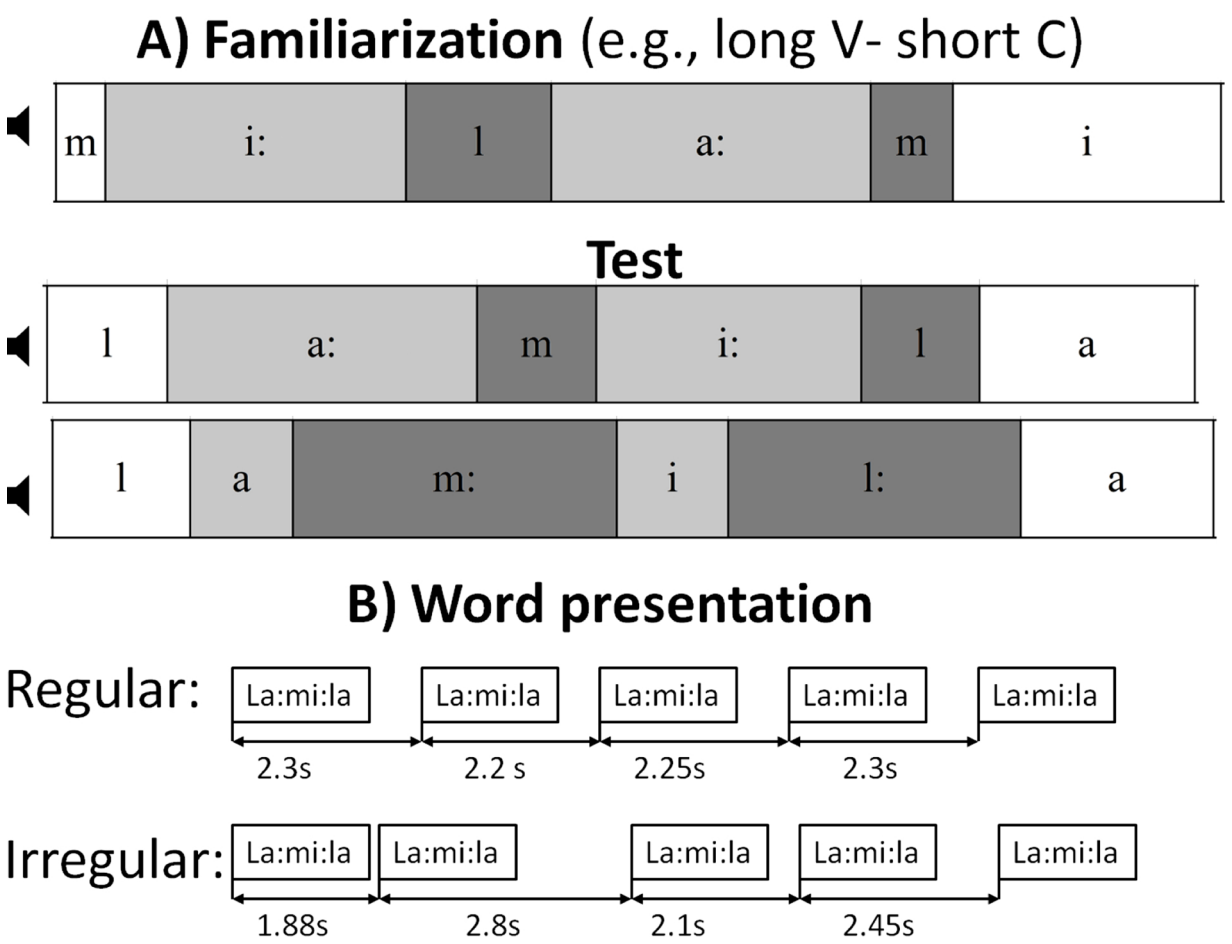

Fig. 1. Structure of an example trial. Panel A) Infants were familiarized with one of the pseudo-word sequences (either short-long or long-short) and tested on the other pseudo-word sequence (short-long AND long-short). Light grey boxes mark vowels while dark grey boxes mark consonants in the critical short-long, and long-short conditions. B) Infants were either assigned to the regular word presentation condition (i.e., inter-word-intervals with slight temporal jitter) or irregular condition (i.e., inter-word-intervals with high temporal jitter) that was applied in both familiarization and test phase.

located directly above that computer monitor. The key press, made by the experimenter, also initiated a timer for the looking-time behaviour of the infant for this particular trial. The stimulus continued playing until the infant looked away (45 degree head turn for 2 s) at which time the experimenter released the key press, terminating the timer for this trial as well as extinguishing the auditory and visual stimulus. The next trial began with Mickey Mouse flashing on the computer monitor on the opposite side (i.e., if the start side was right, the opposite side would be left) of the infant. When the infant focused on the other computer monitor the trial proceeded in an identical manner but presented a stimulus with the opposite durational pattern. This alternation of familiar and novel stimuli continued for 20 trials, such that each stimulus type was presented 10 times in total. The side of first presentation (left or right) and the first stimulus (varying in durational cues) was counterbalanced across participants. The duration of testing was approximately 15 – 20 min.

3. Results

An initial analysis of variance (ANOVA) was conducted to establish that infants had no a priori preference for one of the stimulus words (e.g., milami versus lamila), side of presentation (left versus right), first stimulus presented (short-long versus long-short first), with looking time (in seconds) to familiar stimuli versus novel stimuli as the dependent variable. As infants often habituate to stimuli over the course of several trial presentations, we also controlled for test-session Half (first 10 trials vs. last 10 trials). A significant main effect of familiarity/novelty was found such that infants looked longer to the novel stimulus compared to the familiar stimulus, $F(1, 57) = 4.12$, $p = 0.047$, ES = 0.515, which was also found in the main analysis reported below. No other significant main effects or interactions were found (all $ps > 0.05$). The other variables were collapsed for the subsequent analysis.

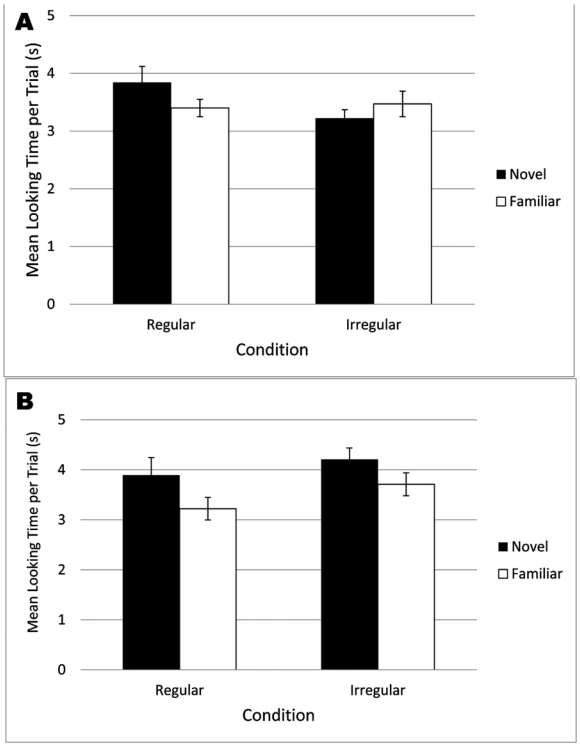

Fig. 2. Mean looking time per trial to novel versus familiar stimuli during the Test Phase. The top panel A) shows looking times in the first 10 trials and the bottom panel B) shows looking times in the last 10 trials. Error bars represent the standard error of the mean.

A 2 × 2 × 2 mixed ANOVA was conducted with Familiarity (familiar versus novel) and test-session Half (trials 1–10 versus trials 11–20) as within-subjects variables, and Temporal regularity (regular versus irregular) as the between-subjects variable. Before running the analysis, we conducted Levene's Test for equality of variances across groups. No significant differences between groups were found,[1] indicating that the assumption of homogeneity was not violated (i.e., the two different groups in the temporal condition (regular/irregular) did not have unequal variances). The results are displayed in Fig. 2 (split by first and second test-session Half, panel A & B, respectively). First, we found a significant main effect of Familiarity, F (1, 57) = 4.12, p = 0.047, ES = 0.52. Overall, infants looked longer towards the stimulus comprising the novel temporal pattern (M = 37.91, SD = 2.08) than the familiar temporal pattern (M = 34.51, SD = 2.80), indicating that they were able to discriminate the contrast. Second, there was also a significant test-session Half x Temporal regularity interaction (F (1, 57) = 5.025, p = 0.029, ES = 0.60). The interaction indicates that infants listened longer to stimuli in the second than in the first half of the Experiment, and that this was particularly the case for the irregular condition. A trend towards significance was found for a main effect of test-session Half (F(1,57) = 3.21, p = 0.079, ES = 0.42), such that infants tended to look longer during the second half of the experiment. No other significant main effects or interactions were found (all ps > 0.05).

As there was a relatively large age range tested in this study (i.e., 6.0–9.1 months), we also included age (in months) as a covariate to account for differences in abilities across the age range of the sample. The covariate analysis showed no main effect of age (F (1, 56) = 1.68, p = 0.20, ES = .25) or any interactions with any other variable with age, (all other ps > 0.05).

4. Discussion

The results provide evidence that English-learning infants between 6–9 months are able to discriminate temporal variations of speech sounds depending on the immediate temporal speech context (i.e., long vowels preceding short consonants, short vowels preceding long consonants at the junction of disyllabic words). Hence, infants are able to discriminate contextual temporal vowel length variations at the same age as they distinguish non-contextual acoustic vowel length variations (Eilers et al., 1984).

Moreover, we show that infants were able to remember, and to a certain extent, abstract these contextual temporal relations and apply them to new instances. Infants can generalize contextual speech information, such as specific co-occurrences of sounds (i.e., phonotactic patterns) from 4 months on, although they start showing biases deriving from ambient language experience in the second half of the first year of life (e.g., Seidl, Cristià, Bernard, & Onishi, 2009). In the present experiment, infants were familiarized with one

[1] Levene's statistic based on the median with adjusted degrees of freedom for each group comparison ranged from 0.004 to .345, with p-values ranging from 0.56 to 0.95

set of words (e.g., /mi:la:mi/), but tested with another set (e.g., /la:mi:la/ /lam:il:a/). Even though the critical vowel-consonant durations (e.g., /i:/-/l/, /a:/-/m/) did not occur in the same syllable positions and varied in stress and intonation between the familiarization and test phases of the experiment, infants discriminated the familiar from the novel pattern.

Unexpectedly, infants' discrimination performance tended to be better in the second half of the Experiment. Typically, many perception studies using the head-turn procedure with this age-group report better discrimination scores during the first half of the experiment and habituation during the second part of the experiment (for a review see Hunter & Ames, 1988). However, when infants are still developing a specific ability, they tend to take time to show an attentional preference, leading to better performance in the second half compared to the first. Likewise, if the task demands are high, infants tend to be slow to perform on a given task (see Gogate & Maganti, 2016, for a general review). For these possible reasons, the infant listeners in the present study may have taken longer (i.e., they needed more trials) to learn the differences between the familiar temporal pattern and the novel stimulus.

We did not find that infants' had greater ease of perceiving the contrast when words were presented at more regular temporal intervals compared to when they were presented at highly irregular intervals. This finding allows us to exclude the hypothesis that it was solely the extraction of a regular "beat" that lead infants to discriminate the contrast. Nevertheless, infants showed longer looking times in the second half of the test when listening to the irregular compared to the regular stimulus. In combination with the near-significant result that discrimination tended to be better in the second than the first half of the experiment, this finding suggests that more irregular word presentation (which may be occurring in more naturalistic speech contexts) is more difficult for infants' to process, thereby delaying their discrimination performance.

Naturalistic speech contexts may have other facilitating characteristics to foster infants' temporal discrimination capacities. For example, adults use an infant-directed speaking style that can foster infants' language development (e.g., Ramírez-Esparza, García-Sierra, & Kuhl, 2014), and potentially impacts vowel discrimination (e.g., Trainor & Desjardins, 2002; Tsao, Liu, & Kuhl, 2004, note that these studies examine spectral characteristics). The amount of exposure to this infant-directed speech style varies widely in the infant and toddler (3–20 months) population of North American households (Bergelson et al., 2019). It is possible that infants more exposed to infant-directed speech could have better or earlier temporal vowel discrimination skills than infants with low exposure. Moreover, as part of infant-directed communication, parents frequently sing infant-directed songs and rhymes to which infants like to listen and attend (Falk, 2011; Trehub & Trainor, 1998; Tsang & Conrad, 2010). These songs display a more regular rhythm than speech (Bergeson & Trehub, 2002), and the rhythm and note-to-syllable mapping in live performances of infant-directed singing render temporal contrasts between long and short vowels and long and short consonants particularly salient (Falk, 2011). Hence, it is an interesting avenue for future research to investigate whether increased exposure to infant-directed speech and singing may foster infants' perception of temporal speech sound patterns.

Finally, by testing a non-native contrast, we conclude that infants' discrimination was based on acoustic auditory cues and not on infants' linguistic experiences. However, we cannot entirely exclude that some infants in our sample may have had incidental exposure to the non-native contrast used in the study in other non-household languages from people outside the home, a factor that was not controlled for. Although highly unlikely (only 1.8 % of the population in BLINDED report a Germanic language as their native language, BLINDED), an incidental exposure may have influenced infant perception of the speech contrasts presented in this study. In a future study, the influence of infants' linguistic environment on temporal discrimination of vowel durations in syllables could be systematically investigated in infants in their second year of life. For example, we would expect that infants learning languages without similar phonological temporal variations should lose the ability to discriminate the unfamiliar contrast compared to infants whose language environment contains temporally conditioned vowel variation (similar to the results of Dietrich et al., 2007). To conclude, the present study shows that infants can discriminate and generalize temporal regularities in speech sound combinations in their first year of life, an early capacity that may help to acquire the complex communicative functions of temporal variation in speech.

CRediT authorship contribution statement

Simone Falk: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **Christine D. Tsang:** Conceptualization, Methodology, Investigation, Data curation, Writing - review & editing.

Declaration of Competing Interest

The authors do not have a conflict of interest to declare

Acknowledgements

This research was supported by a Faculty of Arts and Social Science research grant from Huron University College to CDT. We thank Alyssa Kuiack for her assistance in participant recruitment, data collection and data analysis, and Sabrina Habte for her assistance with participant recruitment.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.infbeh.2020.101475.

References

Bergelson, E., Casillas, M., Soderstrom, M. F., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science, 22*(1), e12724. https://doi.org/10.1111/desc.12724.

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science, 13*, 71–74. https://doi.org/10.1111/1467-9280.00413.

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science, 10*(6), 770–777. https://doi.org/10.1111/j.1467-7687.2007.00635.x.

Chang, H. W., & Trehub, S. E. (1977). Infants perception of temporal grouping in auditory patterns. *Child Development, 48*(4), 1666–1670.

De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493–516. https://doi.org/10.1016/j.wocn.2004.05.002.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences, 104*(41), 16027–16031. https://doi.org/10.1073/pnas.0705270104.

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America, 75*(4), 1231–1238. https://doi.org/10.1121/1.390773.

Elert, C. C. (1964). *Phonologic studies of quantity in swedish.* Uppsala: Almqvist & Wiksell.

Falk, S. (2011). Temporal variability and stability in infant-directed sung speech: Evidence for language-specific patterns. *Language and Speech, 54*(2), 167–180. https://doi.org/10.1177/0023830910397490.

Galle, M. E., & McMurray, B. (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin & Review, 21*(4), 884–906. https://doi.org/10.3758/s13423-013-0569-y.

Gogate, L., & Maganti, M. (2016). The dynamics of infant attention: Implications for crossmodal perception and word-mapping research. *Child Development, 87*(2), 345–364. https://doi.org/10.1111/cdev.12509.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Fledge, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *The Journal of the Acoustical Society of America, 76*(1), 18–26. https://doi.org/10.1121/1.391094.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier, & L. P. Lipsitt (Vol. Eds.), *Advances in infancy research: Vol. 5*, (pp. 69–95). Westport, CT, US: Ablex Publishing.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior & Development, 76*(18), 111–116. https://doi.org/10.1016/0163-6383(95)90012-8.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208–1220. https://doi.org/10.1121/1.380986.

Ko, E. S., Soderstrom, M., & Morgan, J. (2009). Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *The Journal of the Acoustical Society of America, 126*(5), 134–139. https://doi.org/10.1121/1.3239465.

Ladefoged, P. (1996). *The sounds of the world's languages.* Oxford: Blackwell Publishers.

Levinson, C. S., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731. https://doi.org/10.3389/fpsyg.2015.00731.

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4- to 10-month-old infants. *Developmental Psychobiology, 48*(7), 631–632. https://doi.org/10.1002/dev.20140.

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology, 45*(1), 236–247. https://doi.org/10.1037/a0014043.

Nakata, T., & Mitani, C. (2005). Influences of temporal fluctuation on infants 'attention. *Music Perception, 22*(3), 401–409. https://doi.org/10.1525/mp.2005.22.3.401.

Otte, R. A., Winkler, I., Braeken, M. A., Stelenburg, J. J., Van Der Stelt, O., & Van Den Bergh, B. R. (2013). Detecting violations of temporal regularities in waking and sleeping two-month-old infants. *Biological Psychology, 92*(2), 315–322. https://doi.org/10.1016/j.biopsycho.2012.09.009.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*(2), 285–290. https://doi.org/10.3758/bf03213946.

Quené, H., & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*(1), 1–13. https://doi.org/10.1159/000087222.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*, 880–891. https://doi.org/10.1111/desc.12172.

Sato, Y., Kato, M., & Mazuka, R. (2012). Development of single/geminate obstruent discrimination by Japanese infants: Early integration of durational and non-durational cues. *Developmental Psychology, 48*(1), 18–34. https://doi.org/10.1037/a0025528.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology, 46*(1), 106–119. https://doi.org/10.1037/a0016718.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development, 5*(3), 191–202. https://doi.org/10.1080/15475440902754326.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review, 9*(2), 335–340. https://doi.org/10.3758/BF03196290.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science, 7*, 289–296. https://doi.org/10.1111/j.1467-7687.2004.00348.x.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. In C. Rovee-Collier, L. Lipsitt, & H. Hayne (Vol. Eds.), *Advances in infancy research: Vol. 12*, (pp. 43–77). Stamford, CT: Ablex Publishing.

Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior & Development, 33*(1), 96–100. https://doi.org/10.1016/j.infbeh.2009.11.006.

Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants prefer infant-directed song over speech. *Child Development, 88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology, 56*(2), 179–191. https://doi.org/10.1002/dev.21179.

D.Whalen, (1990). Coarticulation is largely planned. Haskins Laboratories Status Report on Speech Research, SR-101 /102, 149–176.

Zheng, X., & Pierrehumbert, J. B. (2010). The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America, 128*(2), 851–859. https://doi.org/10.1121/1.3455796.