Article

Language *and* Speech

When Too Many Vowels Impede Language Processing: An Eye-Tracking Study of Danish-Learning Children

Language *and* Speech
2020, Vol. 63(4) 898–918

DOI: 10.1177/0023830919893390
journals.sagepub.com/home/las
SAGE

Fabio Trecca
School of Communication and Culture, Aarhus University, Denmark

Dorthe Bleses

Anders Højen
TrygFonden's Centre for Child Research, Aarhus University, Denmark

Thomas O Madsen
Department of Language and Communication, University of Southern Denmark, Denmark

Morten H Christiansen
Department of Psychology, Cornell University, NY; Interacting Minds Centre & School of Communication and Culture, Aarhus University, Denmark

Abstract

Research has suggested that Danish-learning children lag behind in early language acquisition. The phenomenon has been attributed to the opaque phonetic structure of Danish, which features an unusually large number of non-consonantal sounds (i.e., vowels and semivowels/glides). The large number of vocalic sounds in speech is thought to provide fewer cues to word segmentation and to make language processing harder, thus hindering the acquisition process. In this study, we explored whether the presence of vocalic sounds at word boundaries impedes real-time speech processing in 24-month-old Danish-learning children, compared to word boundaries that are marked by consonantal sounds. Using eye-tracking, we tested children's real-time comprehension of known consonant-initial and vowel-initial words when presented in either a consonant-final carrier phrase or in a vowel-final carrier phrase, thus resulting in the four boundary types C#C, C#V, V#C, and V#V. Our results showed that the presence of vocalic sounds around a word boundary—especially *before*—impedes processing of Danish child-directed sentences.

Corresponding author:
Fabio Trecca, School of Communication and Culture, Aarhus University, Jens Chr. Skous Vej 2, Aarhus C, DK-8000, Denmark.
Email: fabio@cc.au.dk

Keywords
Language development, speech processing, eye-tracking, consonants, vowels, Danish

Introduction

Children learn their native language in a seemingly effortless way, acquiring an impressive number of words and expressions in a relatively short time span. However, despite general similarities in developmental trends across language groups (e.g., Dale & Goodman, 2005), not all languages are learned at the same rate. Danish-learning children, for example, have been found to lag considerably behind in the early acquisition of receptive vocabulary, compared to children learning a number of other European and non-European languages (Bleses et al., 2008; Bleses, Basbøll, Lum, & Vach, 2011); in the acquisition of past-tense morphology, compared to children learning Norwegian, Swedish, and Icelandic, which are closely related to Danish both genetically and typologically (Bleses, Basbøll, & Vach, 2011); and in the acquisition of other early linguistic milestones, such as labeling and early word combinations, compared to children learning Norwegian and Swedish (Bleses & Trecca, 2016).

Researchers have argued that this delay is possibly due a number of language-inherent characteristics of Danish that may make the speech signal fundamentally harder to process and learn from (e.g., Bleses & Basbøll, 2004; Bleses et al., 2008; Bleses, Basbøll, Lum, et al., 2011; Grønnum, 2003; Trecca, Bleses, Madsen, & Christiansen, 2018; Trecca et al., 2019; Trecca, Tylén, Højen, & Christiansen, under review; see also Gooskens, van Heuven, van Bezooijen, & Pacilly, 2010; Schüppert, Hilton, & Gooskens, 2016). Firstly, spoken Danish has an unusually large inventory of vocalic sounds compared to consonantal sounds. At the phonetic level of analysis, Danish has 29 different monophthongal vowel qualities (16 short and 13 long), 18 falling diphthongs (vowel + [j]), and at least 12 rising diphthongs ([j] + vowel), compared to only 19 consonants (Grønnum, 1998; 2005). This makes Danish one of the languages with the highest ratio of vowels to consonants among the European and North American languages (Bleses, Basbøll, Lum, et al., 2011). Secondly, consonants occurring in unstressed syllables are often realized as semivowels/glides in both casual and distinct speech. Common instances of this pervasive process of consonant weakening in Danish are as follows: (a) the general loss of closure of /b v/, which are then realized as [u̯] (as in *(at) løbe* [løu̯ə], Eng. '(to) run', and in *kniv* [ˈkʰniu̯ˀ], Eng. 'knife') as well as of /g/, which is realized as either [u̯] or [ɪ̯] (as in *(at) koge* [ˈkʰɔːu̯ə], Eng. '(to) boil' or in *(at) bage* [ˈb̥æːɪ̯ə], Eng. '(to) bake'); (b) the pervasive reduction of /d/ to the non-lateral approximant [ð̞], which is a very weak vowel-like sound without friction noise (as in *(at) bade* [ˈb̥æːð̞ə], Eng. '(to) bathe'); and (c) the mandatory realization of the very common plural and present tense suffix /r/ with the schwa-vowel [ɐ] (as in *biler* [ˈb̥iˀlɐ], Eng. 'cars' and *(han/hun/den) spiser* [ˈsb̥iːsɐ], Eng. '(he/she/it) eats').[1] Together, the unusually large vowel inventory and the general weakening of consonants to semivowels/glides result in a speech stream that is distinctively rich in long sequences of phonetic vowels (vocoids) with few or no intervening phonetic consonants (contoids). This is the case in a common expression like *her er jeg* [ˈhɑ æ ˈjɑj], Eng. "here I am," in which the initial [h] is followed by a string of five adjacent vocoids, straddling the boundaries between three words, with no clear amplitude or sound quality discontinuities (see Trecca et al., 2018, 2019).

Together, these phonetic peculiarities have been hypothesized to make Danish speech especially hard for children to process in real time (e.g., Bleses & Basbøll, 2004; Trecca et al., under review). This suggestion builds on at least three arguments: Firstly, consonants tend to be perceptually more salient than vowels (e.g., Lieberman, Harris, Hoffman, & Griffith, 1957), and consequently they seem to provide better cues to word boundaries. For instance, English-learning children at the age of 13.5 months were shown to successfully segment consonant-initial verbs with a weak–strong stress pattern (e.g., perMIT) from continuous speech, whereas they failed to segment vowel-initial

verbs with the same stress pattern (e.g., imPORT) until 16.5 months of age (Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel, & Jusczyk, 2005). Similarly, English consonant-initial nouns were found to be readily segmented from fluent speech at 8 months of age, while vowel-initial nouns were not segmented correctly until 16 months (Mattys & Jusczyk, 2001). Children younger than 16 months of age have been shown to successfully segment vowel-initial nouns, but only when these occur in highly salient positions (Kim & Sundara, 2015; Seidl & Johnson, 2008).

Secondly, consonants are believed to carry more lexical information than vowels at the phonological level, therefore playing a primary role in word processing, whereas vowels may be more useful at encoding syntactic and prosodic information (e.g., Nespor, Peña, & Mehler, 2003). Evidence for such consonantal bias in lexical processing has been found in young children from different language groups (see Nazzi, Poltrock, & Von Holzen, 2016, for a review). For instance, French-learning children as young as 11 months were shown to tolerate vowel mispronunciations of familiar words to a higher degree than consonant mispronunciations in a word recognition task, suggesting that the former are less disruptive for lexical identification, at least in French (Poltrock & Nazzi, 2015). At 16 months, French-learning children can successfully learn nonsense words that contrast by a consonant (e.g., /pyf/ vs. /tyf/), but not those that contrast by a vowel (e.g., /pœs/ vs. /pos/) (Havy & Nazzi, 2009; see also Nazzi, 2005, for similar outcomes with 20-month-olds). Similarly, Italian-learning children as young as 12 months of age were found to rely more on consonants than vowels when establishing novel phonological representations of nonsense words (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011), and Canadian French-learning 20-month-olds were shown to be significantly better at learning consonant contrasts (e.g., *oupsa* vs. *outsa*) than vowel contrasts (e.g., *opsi* vs. *eupsi*). Even in adults, consonants have been found to facilitate lexical access in a lexical decision task to a higher extent than vowels, in both speakers of French and English (Delle Luche et al., 2014; see also New, Araújo, & Nazzi, 2008).

Interestingly, both French- and Italian-learning children seem to start out by favoring vocalic information in their first months of life, before switching to weighting consonants more at a later age (e.g., Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012; Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015; Nishibayashi & Nazzi, 2016). This raises the question as to whether a bias for consonants may not be universal but acquired over time, as the child gets attuned to his or her ambient language (see Nazzi et al., 2016). For instance, Danish-learning children show an uncommon bias for vowels at 20 months, suggesting that a lexical processing bias for consonants versus vowels may be language specific, rather than universal (Højen & Nazzi, 2016). In this word learning study, Danish-learning children were able to learn phonetically similar non-words that contrasted on vowels (e.g., /dyl/ vs. /dul/), but not those that contrasted on consonants (e.g., /fan/ vs. /san/). This finding has been taken to suggest that Danish-learning children, compared to children learning less vowel-heavy languages, may adapt early on to the highly vocalic nature of their input.

Thirdly, the highly vocalic sound structure of Danish may make other important suprasegmental cues (e.g., prosody, phonotactics) and distributional cues (e.g., segment counts, syllable counts, transitional probabilities) to syllable and word boundaries less prominent for the novice listener to pick up on (e.g., Stokes, Bleses, Basbøll, & Lambertsen, 2012). Consider, for instance, the regular past-tense Danish suffix *-ede*, as in *(at) bade* → *badede* ([ˈbæːðəðə], Eng. "bathe, bathed"), which is realized in fluent speech as [ˈbæːðː], with the last two syllables being reduced to one long [ð]-sound with no prosodic nor distributional cues to syllabic structure (Bleses, Basbøll, & Vach, 2011). This phonetic (in contrast to phonological) loss of one or more syllables may have negative consequences for the extent to which Danish-learning children can rely on statistical learning mechanisms (such as computing transitional probabilities across adjacent and non-adjacent syllables) in order to process the speech input (cf. Trecca et al., 2019).

Therefore, Danish speech seems to contain intrinsically fewer cues to word boundaries and syllable structure than other closely related languages. Danish-learning children are very likely to encounter morphological and lexical boundaries that are embedded within sequences of vocoids: for instance, 23% of all vocoid–vocoid (V#V) segment pairs in Danish child-directed speech straddle a word boundary, whereas this is only true of 9% of V#V segment pairs in English child-directed speech (Trecca et al., 2019). For this reason, the sound structure of Danish has been often held accountable for the delay observed in the early acquisition of Danish, based on the assumption that an opaque speech signal requires more time and cognitive effort to process and learn from (e.g., Bleses et al., 2008; Bleses, Basbøll, Lum, et al., 2011). At the same time, Danish-learning children do show an uncommon vowel-processing bias when learning new words, that is, a better ability for processing paradigmatic vowel distinctions compared to consonant distinctions. This suggests a vocalic processing bias shaped by the phonetic peculiarities of their ambient language during early childhood (Højen & Nazzi, 2016). Such an adaptation to the vocalic nature of Danish may then be expected to facilitate—rather than impede—processing, potentially mitigating some of the challenges associated with processing vocalic sounds described above.

Motivated by these previous findings, we were interested in testing to what extent the presence of vocoids in fluent speech would negatively affect the comprehension of familiar words in a sample of Danish-learning two-year-old children. To test this question empirically, we devised an eye-tracking procedure based on the looking-while-listening (LWL) paradigm (Fernald, Zangl, Portillo, & Marchman, 2008). In this paradigm, children are presented with pairs of pictures of familiar objects on a screen, while one of the two objects is named by a recorded voice off screen (e.g., "Look at the *car*!"). Patterns of looking at the target object are recorded via eye-tracking and are taken as a measure of real-time processing skills. In the present study, we used LWL to present children with a number of familiar consonant-initial or vowel-initial words, which were embedded in two common Danish child-directed carrier phrases ending either in a phonetic consonant or a phonetic vowel. This resulted in a 2×2 experimental design (carrier phrase final segment × target word initial segment), yielding four conditions in which the boundary between the carrier phrase and the target word was either delimited by two consonantal sounds (C#C), by one consonantal and one vocalic sound (C#V or V#C), or by two vocalic sounds only (V#V). With this design, we were specifically interested in answering three questions, as follows. (1) Does the presence of phonetic vowels at word boundary (V#C/C#V/C#C vs. V#V) impede real-time language processing—measured as the recognition of a target word in sentence-final position—in Danish 24-month-olds? (2) If so, is this effect on processing differential, depending on whether the phonetic vowels occur before, after, or on both sides of the word boundary (i.e., C#V vs. V#C vs. C#C)? (3) Is the number of contoids at the word boundary directly proportional to processability (C#C > C#V/V#C and C#V/V#C > V#V)?

We set out to answer these questions by measuring differences in the children's looking patterns at pairs of images in response to the four types of speech stimuli. This manipulation consisted of a total of eight trials, which we will refer to as Test trials. However, given that the carrier phrases used in our Test phase deviated from those that are more commonly used in LWL experiments, we also used additional trials with sentence stimuli that were syntactically, semantically, and pragmatically closer to those used in previous studies of, for instance, English-learning children (e.g., Fernald, Thorpe, & Marchman, 2010) for replicability purposes. We will refer to these trials as Baseline trials.

2 Method

2.1 Participants

Our participants were 43 24-month-olds from monolingual native Danish-speaking families recruited from the Odense Child Cohort (Kyhl et al., 2015) in the Odense area of Denmark. To be

included in the final analyses, each child had to contribute at least 65% of gaze data per trial in at least half of the trials (i.e., the amount of data loss per child was not allowed to exceed 35% of all the possible data points in a trial, in 11 out of the 22 total trials). Sixteen children were excluded from the analyses for not satisfying this criterion. Data from five children were collected, but not included in the analyses due to the children not completing the procedure ($n = 2$), experimenter error ($n = 2$), or the child being exposed to languages other than Danish at home ($n = 1$). The final sample consisted therefore of 22 children ($M_{age} = 23.3$, SD = 0.92, range = 22–25, six girls, 16 boys).

2.2 Speech stimuli

The stimulus sentences consisted of two carrier phrases followed by one of four target words. In Baseline trials, the carrier phrases were *Kan du se . . .[-en]?* ("Can you see . . .[-the]?"; notice the postponed definitive article *-en* in Danish) and *Kan du se en . . .?* ("Can you see a . . .?"). Despite what the orthography suggests, both carrier phrases have CV#CV#CVC phonetic structure and are therefore contoid-final (the last C in the first carrier phrase being the Danish *stød*, which in our stimuli is realized phonetically as a glottal stop with a mean silent interval of 165 ms). In Test trials (in which our phonetic manipulation was implemented), the carrier phrases were the contoid-final *Find . . .[-en]!* (Eng. "Find . . .[-the]!") with CVCC structure (where the last C is also glottal stop with a mean silent interval of 181 ms), and the vocoid-final carrier phrase *Her er . . .[-en]!* (Eng. "Here's . . .[-the]!", where the two /r/ are realized as [ɐ̯]), with CVV#V structure. The target words were either consonant-initial (*bamse*, Eng. "teddy bear"; *bil*, Eng. "car") or vowel-initial (*abe*, Eng. "monkey"; *and*, Eng. "duck"). The combinations of carrier phrases and target words resulted in eight possible stimulus sentences in each trial type. In Baseline trials, the resulting stimulus sentences were as follows: 2 × C_1#C (*Kan du se bamsen?* and *Kan du se bilen*), 2 × C_1#V (*Kan du se aben?* and *Kan du se anden?*), 2 × C_2#C (*Kan du se en bamse?* and *Kan du se en bil?*), and 2 × C_2#V (*Kan du se en abe?* and *Kan du se en and?*). In Test trials, the resulting sentences were as follows: 2 × C#C (*Find bamsen!* and *Find bilen!*), 2 × C#V (*Find aben!* and *Find anden!*), 2 × V#C (*Her er bamsen!* and *Her er bilen!*), and 2 × V#V (*Her er aben!* and *Her er anden!*). Phonetic transcriptions and mean duration of the speech stimuli used in the experiment are reported in Table 1.

The six additional words *baby* (baby), *bog* (book), *bold* (ball), *hund* (dog), *kat* (cat), and *ko* (cow) were used as target words in familiarization trials at the beginning of the procedure, as well as in filler trials that were interspersed throughout the procedure (see Section 2.4). Each sentence was followed, 800 ms after the target word offset, by one of three attention-getting phrases (*Kan du finde den?*, "Can you find it?"; *Kan du se den?*, "Can you see it?"; *Kig på den!*, "Look at it!"). Two infant-directed reinforcement sentences (*Kunne du lide billederne? Her kommer flere!*, "Did you like the pictures? Here are some more!" and *Det var flot klaret!*, "Well done!") were played respectively halfway through the procedure and at the very end, accompanied by colorful drawings of cartoon characters. All speech stimuli were recorded in child-directed form by a female native speaker of Danish.

The 10 target words (four in test stimuli + six in familiarization/filler stimuli) were chosen using vocabulary norms based on the Danish adaptation of the MacArthur-Bates Communicative Development Inventories (MB-CDI) (Bleses, Vach, Wehberg, Faber, & Madsen, 2007), which were retrieved via the web-based cross-linguistic MB-CDI lexical norms database CLEX (Jørgensen, Dale, Bleses, & Fenson, 2010). In the database, the four test stimuli target words *bamse, bil, abe*, and *and* are produced by 81%, 93%, 67%, and 67% of children, respectively. However, all children in our study knew all four words, both receptively and productively; this was established by administering a short vocabulary checklist to the parents at the time of the experiment. The six familiarization/filler

Table 1. Summary of the speech stimuli.

	Phonetic structure	Danish stimuli	Phonetic realization[a]	English translation	Mean duration
Carrier phrases	*In Baseline trials*:				
	Neutral 1 (CV#CV#CVC)[b]	*Kan du se . . . [-en]?*	[ʔʔæ dʔu ʔseʔʔʔ] [ʔʔædʔuʔseʔʔʔn]	*Can you see the . . . ?*	535 ms, *SD* = 52.5
	Neutral 2 (CV#CV#CVC)	*Kan du se en . . . ?*		*Can you see a . . . ?*	590 ms, *SD* = 22.9
	In Test trials:				
	Contoid-final (CVCC)[b]	*Find . . .[-en]!*	[ʔfeʔnˀ]	*Find . . . !*	279 ms, *SD* = 28.3
		Her er . . .[-en]!	[ʔheʔʔ æ]	*Here is . . . !*	
	Vocoid-final (CVV#V)				276 ms, *SD* = 41.6
Target words	Consonant-initial	*bamse* / *bamsen*	[ʔbʔʔmsʔ] / [ʔbʔʔmsnʔ]	*(the) teddy bear*	471 ms, *SD* = 11.3
		bil / *bilen*	[ʔbʔiʔˀl] / [ʔbʔiʔˀlnʔ]	*(the) car*	462 ms, *SD* = 39.5
	Vowel-initial	*abe* / *aben*	[ʔʔʔʔbʔʔ] / [ʔʔʔʔbʔmʔ]	*(the) monkey (the) duck*	388 ms, *SD* = 39.9
		and / *anden*	[ʔaʔnˀnʔ] / [ʔaʔnˀnʔ]		476 ms, *SD* = 45.2

[a]Non-normalized International Phonetic Alphabet (IPA) transcription based on Basbøll (2005).
[b]The last C is the Danish *stød*, here phonetically realized as a glottal stop with a mean silent interval of 165 ms in the Neutral 1 carrier phrase, and of 181 ms in the contoid-final carrier phrase.

stimuli target words, *baby, bog, bold, hund, kat*, and *ko*, were known productively by 80%, 87%, 91%, 88%, 82%, and 84% of children, respectively.

2.3 Visual stimuli

In each trial, the children saw one target image and one distractor image, measuring 800×800 pixels each, one on each side of the screen. The images were photographic depictions of the different target objects. Two different picture tokens were used for each word to counter the effects of habituation and natural preferences for specific images. All pictures served both as target and distractor across trials. Trial sequence and location of presentation for images were quasi-randomized across participants, with each participant being randomly assigned to one of four possible trial sequences (see Supplemental Material for details).

2.4 Procedure

The procedure consisted of 22 trials, which unfolded as outlined Table 2. Firstly, the children were presented with three Familiarization trials intended to make them acquainted and comfortable with the procedure. In these trials, the children were presented with the six familiarization/filler words described in Section 2.2. Three of the words served as targets, and the remaining three served as distractors. After the familiarization phase, the children were presented with eight Baseline trials and eight Test trials. In the Test trials, the children were presented with our contoid-final and vocoid-final test sentence stimuli. In both types of trials, the two consonant-initial words *bamse* and *bil* and the two vowel-initial target words *abe* and *and* served as both targets and distractors in

Table 2. Structure of the procedure.

Trial type	Number of trials
Familiarization	3
Baseline	4
Filler	1
Baseline	4
Filler	1
Test	4
Filler	1
Test	4
Total	22

Note. The order of presentation of Baseline trials and Test trials was quasi-randomized across participants.

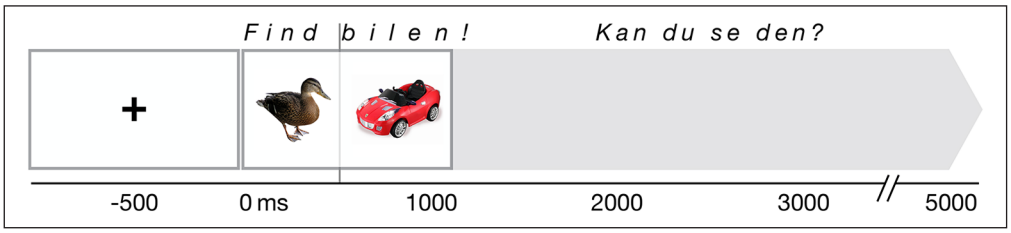

Figure 1. Schematic representation of the trial structure. Each trial started with a 500 ms attention grabber (a static picture of a toy), followed by the onset of the visual stimuli together with the speech stimuli. The vertical line shows the approximate onset of the target word. An attention-grabbing sentence followed after 800 ms from the end of the target word.

different trials. The order of presentation of the Baseline trials and Test trials was quasi-randomized across children. Lastly, three Filler trials—equivalent to the Familiar trials described above—were interspersed throughout and between Baseline and Test trials to maintain the children's attention and counter habituation. The pairings between target and distractor in each trial were pseudorandomized into four possible presentation orders (see Supplemental Material) to which each child was randomly assigned. Targets and distractors were not presented in yoked pairs, but a consonant-initial word (i.e., *bamse* or *bil*) was always paired with a vowel-initial word (i.e., *abe* or *and*) in each trial.

The structure of the individual trials is laid out in Figure 1. Each trial began with a 500 ms fixation screen in which a colorful picture of a toy, randomly picked among three possible pictures, was shown in the center of the screen. The fixation screen was then followed by the presentation of the two pictures along with the speech stimulus.[2] Target word onset happened on average at the 592 ms timestamp in Baseline trials and at the 287 ms timestamp in Test trials. One attention-getting phrase was played 800 ms after target word offset, after which the two images remained on screen until the end of the trial. Each trial lasted for a total of 5000 ms, excluding fixation.

2.5 Apparatus and data analysis

The experiment was conducted in a soundproof room at the eLab at the University of Southern Denmark (Odense, Denmark), using a 50" plasma screen and two forward-facing loudspeakers.

Gaze data were collected with a Tobii X120 eye-tracker at a sampling rate of 60 Hz, which was calibrated using an infant-friendly five-point calibration procedure. The children sat on the parent's lap at approximately 60 cm from the eye-tracker and approximately 140 cm from the screen. Parents listened to superimposed music and speech through over-ear headphones and were instructed not to interact with the child.

Data analyses and statistical modeling were carried out in R version 3.6.0 (R Core Team, 2019) with the help of the eyetrackingR package version 0.1.8 (Dink & Ferguson, 2015). We assessed the children's proportional looks at target image and shifts from distractor to target image using Bayesian multilevel regressions. The models were fitted in R with the help of the brms package version 2.9.0 (Bürkner, 2017) using four parallel Markov Chain Monte Carlo sampling chains with 4000 iterations each, an adapt_delta parameter of 0.9, and a max_treedepth parameter of 20. Models M1 and M2 (see Section 3) were linear models of aggregated binomial outcomes that were modeled according to a zero-one inflated beta distribution with a logit link. Models M3 and M4 were logistic models of binary outcomes (1,0) that were modeled according to a Bernoulli distribution, also with a logit link. All models had weakly conservative priors for intercept (normal[$\mu = 0, \sigma = 1$]), beta estimates (normal[$\mu = 0, \sigma = 1$]), and SDs of random effects (normal[$\mu = 0, \sigma = .2$]); models M2 and M4 also had a prior for the correlation coefficients of random effects (lkj[$\eta = 5$]). Specifications for the individual models are reported in the Section 3, and details on the model fits are available as Supplemental Material. In line with previous studies (e.g., Fernald & Marchman, 2012; Fernald, Marchman, & Weisleder, 2013), we constrained our analysis on a 1500-ms time window starting at 300 ms from target onset in order to allow for the time needed to program an eye movement. Two Areas of Interest (AOIs) measuring 800×800 pixels each circumscribed the two pictures on the screen. Non-AOI looks were not included in the analysis. Raw proportional looking data served as our dependent measure in the analyses reported in this paper; repeating the analyses using either arcsine-root transformed or logit-adjusted proportion data produced equivalent results.

3 Results

Figure 2 shows the children's mean proportional looking times at the target image (i.e., the total number of looks at the target object per child divided by the total looks at both target or distractor object, within the time window of analysis) averaged over the 1500-ms time window. An intercept-only multilevel regression model, M1: *Proportional Looks ~ 1 + (1 | Child) + (1 | Item)*, showed that the total proportional looks were credibly above chance level in all conditions in both Baseline trials ($\beta = 0.25$, SD = 0.12, CI[95%][3] = 0.02,0.47, BF[Intercept > 0] > 100, Post.Prob. = 1) and Test trials ($\beta = 0.48$, SD = 0.11, CI[95%] = 0.26,0.72, BF[Intercept > 0] > 100, Post.Prob. = 1), suggesting that the children were responding to the speech stimuli. In order to investigate how looking patterns varied in our different experimental conditions, we started out by quantifying proportional looks to target separately for the two Carrier phrase types and the two Target word types. For the Baseline trials, we fitted a mixed-effects model, M2a: *Looks at Target ~ Target Word + (1 + Target Word | Child) + (1 | Item)*, predicting the proportion of fixations to the target picture in the 1500 ms window as a function of the target word being either consonant-initial or vowel-initial. The model had therefore a population-level effect term for Target Word type (consonant-initial, vowel-initial). No effect term for Carrier Phrase was included, since the two carrier phrases used in Baseline trials did not differ in terms of phonetic or pragmatic characteristics, and since we had no hypothesis about an effect of Carrier Phrase; model comparison based on Bayesian leave-one-out-cross-validation (LOO-CV; Vehtari, Gelman, & Gabry, 2017) confirmed that adding a term for Carrier Phrase worsened the fit of the model (model without Carrier Phrase: $\Delta LOOIC = 0, \Delta SE = 0$; model with Carrier Phrase: $\Delta LOOIC = -1, \Delta SE = 1.4$). In addition to the population-level term, the

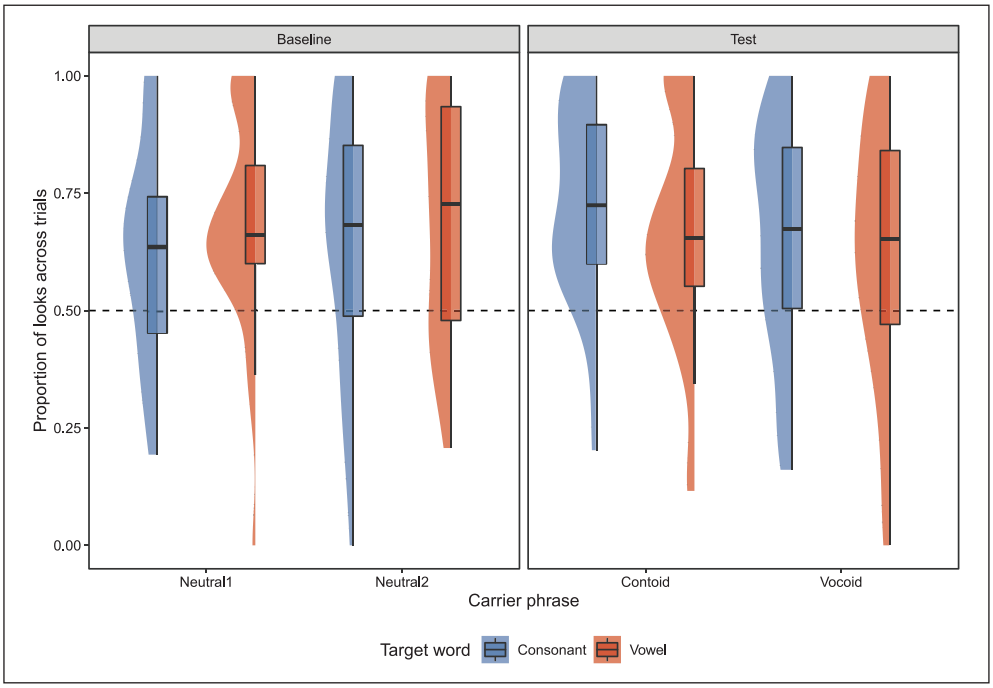

Figure 2. Boxplots and density distributions for proportional looks at target, collapsed across the 1500-ms time window, and divided by trial type (Baseline vs. Test), Carrier Phrase, and Target Word.

model had random intercepts for the individual children with random slopes for the population-level term for each child and random intercepts for the specific speech items. The output of the model for population-level effects is presented in Table 3 (data for group-level effects are available in the Supplemental Material). The model showed a slightly greater number of looks at vowel-initial target words, although we did not find credible evidence for this numerical difference (as suggested by the 95% credible intervals including zero, and by a very strong Bayes Factor[4] in favor of the null hypothesis: BF[TargetWordVowel = 0] = 54.4, Post.Prob. = 0.98).

We also fitted a mixed-effects model to our data from the Test trials, this time predicting the probability of fixating the target picture in the 1500-ms window as a function of the presence of contoids on either versus both sides of the target word boundary. Compared to the model for Baseline trials, this model had therefore population-level terms for both Carrier Phrase type (contoid-final, vocoid-final) and Target Word type (consonant-initial, vowel-initial), which were also added as random slopes for the individual children, M2b: *Looks At Target ~ Carrier Phrase × Target Word + (1 + Carrier Phrase × Target Word | Child) + (1 | Item)*. The output of the model (Table 3) suggested that, when collapsing across the time window, the children looked numerically more at the target objects when their labels were consonant-initial words than vowel-initial words ($\beta = -0.19$, SD = 0.31, CI[95%] = −0.82,0.41), as well as when the carrier phrase was contoid-final rather than vocoid-final ($\beta = -0.32$, SD = 0.28, CI[95%] = −0.86,0.26). The model reported substantial evidence for the effect of Carrier Phrase (BF[CarrierPhraseVocoid < 0] = 7.47, Post.Prob. = 0.88), but not for the effect of Target Word (BF[TargetWordVowel < 0] = 2.89, Post.Prob. = 0.74), nor for the interaction effect between Carrier Phrase and Target Word ($\beta = 0.26$, SD = 0.43, CI[95%] = −0.55,1.16, BF[CarrierPhraseVocoid:TargetWordVowel = 0] = 21.06, Post.Prob. = 0.95).

Table 3. Statistical models of total proportional looks at target in the 1500- ms time window.

	Model M2a: Baseline trials			Model M2b: Test trials		
	β	SD	CI [95%]	β	SD	CI [95%]
Intercept	0.31	0.11	0.09,0.52	0.66	0.22	0.23,1.10
Carrier Phrase: Vocoid	–	–	–	−0.32	0.28	−0.86,0.26
Target Word: Vowel	0.09	0.15	−0.19,0.39	−0.19	0.31	−0.82,0.41
Carrier Phrase: Vocoid × Target Word: Vowel	–	–	–	0.26	0.43	−0.55,1.16

Note. Model M2a: *Looks at Target ~ Target Word + (1 + Target Word | Child) + (1 | Item)*. Model M2b: *Looks at Target ~ Carrier Phrase × Target Word + (1 + Target Word | Child) + (1 | Item)*. Values in the table are on the log-odds scale. Plots of the marginal effects of the two models are available as Supplemental Material.

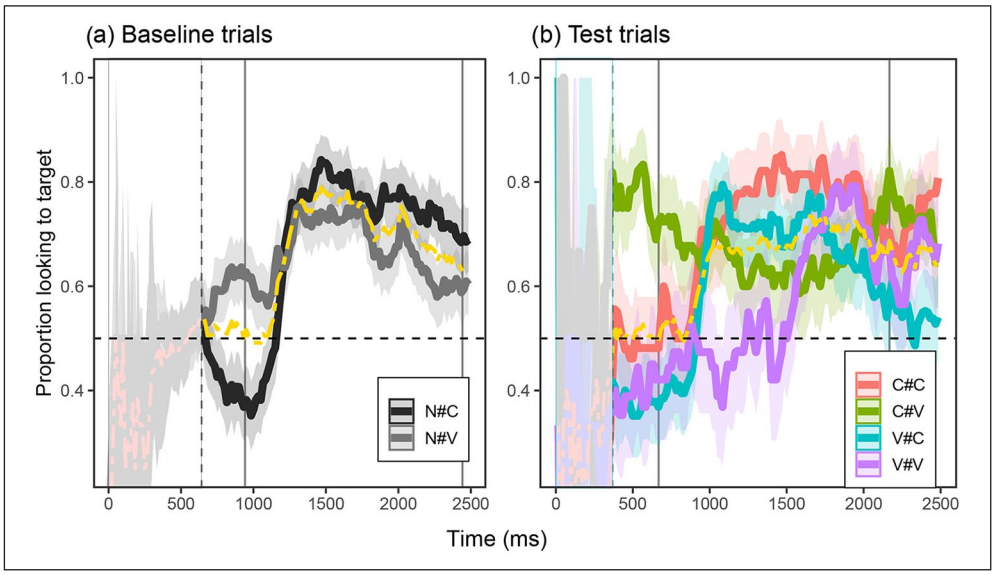

Figure 3. Proportional looks at target as they unfolded throughout the 1500-ms time window for Baseline trials (a) and Test trials (b). The horizontal dashed line indicates the chance level of 0.5. The two vertical lines represent respectively the average beginning and end of the time window over which our analyses were carried out. The dashed yellow lines represent the average of the different gaze patterns. (Color online only.)

We then looked at how gaze patterns at target versus distractor images changed across time. Figure 3 illustrates the overall unfolding of looking patterns over the time of a trial for Baseline trials and Test trials, plotted separately for the different carrier phrases and target words. In Baseline trials, the proportional looks to target were as expected at chance level in the early portion of the trials, with a general shift in gaze toward the target image halfway through the time window (the yellow dotted line shows the average looking pattern when collapsing across the two Target Word types). A bootstrapped cluster-based permutation analysis (Maris & Oostenveld, 2007) based on 2000 permutations was used to compare the average looking pattern to chance level (to which we added a noise term) and showed that mean proportional looks were positively different from chance

in the time window 1166.9–2217.11 ms ($t = -293.6$, $p < 0.001$; further details on the bootstrapped cluster-based permutation analysis can be found in the Supplemental Material). To get a more specific view of how looking patterns at target versus distractor changed across trials in response to the speech stimuli, we also ran a permutation analysis to test whether the gaze patterns were in relation to consonant-initial versus vowel-initial target words, although this did not find any significant areas of divergence (see Supplemental Material). We then modeled the looking patterns in a growth curve analysis (Mirman, Dixon, & Magnuson, 2008). This was done by predicting changes in proportional looking as caused by the different stimuli in relation to linear, quadratic, and cubic time (first-, second-, and third-order orthogonal polynomial terms). The three time-terms allowed us respectively to capture (a) the overall linear growth of the gaze patterns, (b) their general degree of curvature across time, and (c) the sigmoidal inflections in the curves, that is, fluctuations in steepness/shallowness of the looking curves across time. We ran a logistic multilevel model on the data, again predicting the proportion of fixations to the target picture as a function of the two target words (consonant-initial vs. vowel-initial, i.e., C#C vs. C#V), but with the addition of time-terms for the growth curve analysis, M3a: *Looks At Target ~ Target Word × (Linear Time + Quadratic Time + Cubic Time) + (1 + Target Word | Child) + (1 | Item)*. The model fit revealed a significant effect of Cubic Time ($\beta = 25.59$, SD = 3.37, CI[95%] = 18.88,32.22, BF[Cubic Time $> 0] > 100$, Post.Prob. = 1), confirming that the rate of change follows a sigmoidal pattern across the trial. The model fit also revealed a very strong disadvantage for vowel-initial target words ($\beta = -1.34$, SD = 0.50, CI[95%] = −2.3,−0.4, BF[Target Word $< 0] > 100$, Post.Prob. = 1) and a credible interaction between Target Word and Quadratic Time ($\beta = 21.97$, SD = 6.40, CI[95%] = 9.6,34.5, BF[Target Word $< 0] > 100$, Post.Prob. = 1), suggesting that looks at vowel-initial target objects had a steeper and steadier rise throughout the first half of the time window than looks at consonant-initial target objects (see Table 4).

In Test trials, the mean proportional looks at target followed a very similar pattern, starting out at chance level and increasing to a mean proportion of around 70% halfway through the window at approximately 1 s from average target word onset. Permutation analysis showed that mean proportional looks were significantly different from chance in the time window 966.86–2217.11 ms ($t = -278.2$, $p < 0.001$). When comparing across conditions, the permutation analysis revealed a significant divergence between the C#C and V#V condition in the time window 1183.57–1533.64 ms ($t = 61.06$, $p = 0.016$), suggesting that proportional looks at target in the V#V condition become on average significantly different from chance 351 ms later than looks in the C#C condition. As in the Baseline trials, we ran a growth curve analysis to determine whether the different Carrier Phrase × Target Word combinations at target word boundary (i.e., C#C, C#V, V#C, V#V) affected looking patterns, M3b: *Looks At Target ~ Carrier Phrase × Target Word × (Linear Time + Quadratic Time + Cubic Time) + (1 + Carrier Phrase × Target Word | Child) + (1 | Item)*. The model showed very strong evidence in favor of a main effect of Cubic Time ($\beta = 22.88$, SD = 4.88, CI[95%] = 13.1,32.3, BF[CubicTime $> 0] > 100$, Post.Prob. = 1) and—as was the case for model M2b—very strong evidence for a main effect of Carrier Phrase, with less proportional looking at target associated with vocoid-final carrier phrases ($\beta = -1.4$, SD = 0.65, CI[95%] = −2.7,−0.13, BF[CarrierPhraseVocoid $< 0] = 62.49$, Post.Prob. = 0.98). Moreover, we found decisive evidence of a negative interaction between Target Word and Cubic Time ($\beta = -16.42$, SD = 5.73, CI[95%] = −27.6,−5.2, BF[TargetWordVowel:CubicTime $< 0] > 100$, Post.Prob. = 1), which suggests that looks at target in relation to vowel-initial target words have a slower sinusoidal rise than those in consonant-initial target words (Figure 4). Thus, gaze shifting toward vowel-initial target words happens more slowly and less robustly than for consonant-initial target words, independently of whether the carrier phrases is contoid- or vocoid-final.

Table 4. Growth curve analyses.

	Model M3a: Baseline trials			Model M3b: Test trials		
	β	SD	CI [95%]	β	SD	CI [95%]
Intercept	−0.27	0.45	−1.1,0.6	−0.01	0.57	−1.1,1.1
Carrier Phrase: Vocoid	–	–	–	−1.4	0.65	−2.7,0.1
Target Word: Vowel	−1.34	0.5	−2.3,−0.4	0.47	0.69	−0.9,1.8
Linear Time	−9.84	9.16	−27.7,8.1	−8.17	9.33	−26.3,10.4
Quadratic Time	6.61	5.75	−4.6,17.8	10.04	6.69	−3.7,23.6
Cubic Time	25.59	3.37	18.9,32.2	22.88	4.88	13.1,32.3
Carrier Phrase: Vocoid × Target Word: Vowel	–	–	–	0.9	0.81	−0.7,2.5
Carrier Phrase: Vocoid × Linear Time	–	–	–	−9.73	9.31	−27.6,8.6
Carrier Phrase: Vocoid × Quadratic Time	–	–	–	15.45	7.43	1.3,29.9
Carrier Phrase: Vocoid × Cubic Time	–	–	–	10.42	5.53	−0.2,21.4
Target word: Vowel × Linear Time	−13.29	9.15	−31.4,4.3	4.57	9.64	−14.4,23.3
Target word: Vowel × Quadratic Time	21.97	6.4	9.6,34.5	−4.85	7.68	−19.5,10.1
Target word: Vowel × Cubic Time	7.51	4.07	−0.4,15.6	−16.4	5.73	−27.6,−5.2
Carrier Phrase: Vocoid × Target Word: Vowel × Linear Time	–	–	–	4.9	9.61	−13.9,23.8
Carrier Phrase: Vocoid × Target Word: Vowel × Quadratic Time	–	–	–	−10.8	8.09	−26.5,5.1
Carrier Phrase: Vocoid × Target Word: Vowel × Cubic Time	–	–	–	6.95	6.36	−5.5,19.7

Note. Model M3a: *Looks at Target ~ Target Word × (Linear Time + Quadratic Time + Cubic Time) + (1 + Target Word | Child) + (1 | Item)*. Model M3b: *Looks at Target ~ Target Word × (Linear Time + Quadratic Time + Cubic Time) + (1 + Carrier Phrase × Target Word | Child) + (1 | Item)*. Values in the table are on the log-odds scale. Plots of the marginal effects of the two models are available as Supplemental Material.

Figure 3 revealed that proportional looks at target in relation to the C#V boundary type in the Test phase had a peculiar pattern compared to the other boundary types. In order to control for possible pre-naming preferences for specific pictures that may have led to this idiosyncratic pattern, we looked again at the data after having recoded proportional looking depending on whether the children made a correct shift (i.e., a shift to the target image in trials in which the child started out by looking at the distractor image at the onset of the target word) or an incorrect shift (i.e., shifts to the distractor image in trials in which the child started out by looking at the target image at the onset of the target word). These data are plotted in Figure 5. The solid lines show correct shifts from distractor to target across the time window for the four different Carrier Phrase × Target Word boundary types. The four lines show a similar gaze shift pattern for the four boundary types, with the rate being highest for the C#C boundary type and progressively lower for C#V, V#C, and

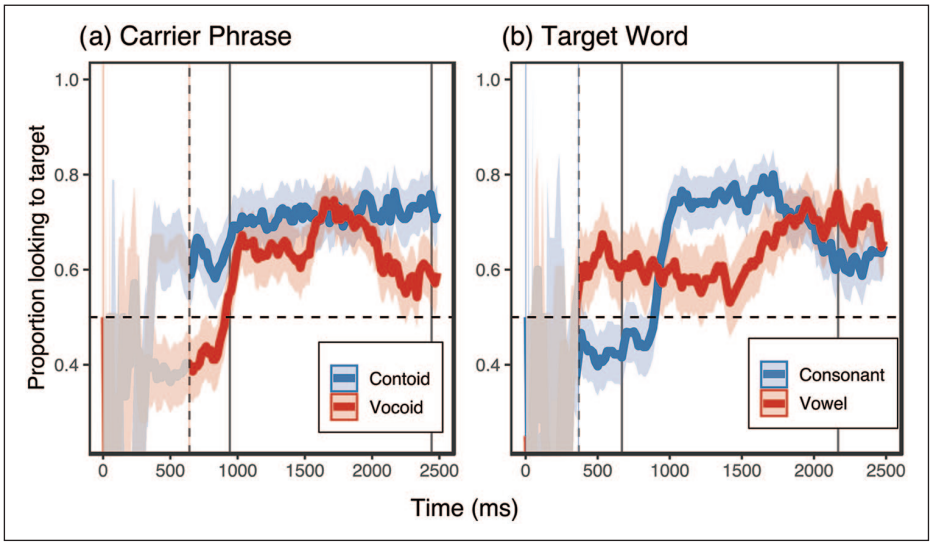

Figure 4. Proportional looks at target as they unfolded throughout the 1500-ms time window for contoid-final versus vocoid-final carrier phrases (a) and for consonant-initial versus vowel-initial target words (b) in Test trials. The horizontal dashed line indicates the chance level of 0.5. The two vertical lines represent respectively the average beginning and end of the time window over which our analyses were carried out.

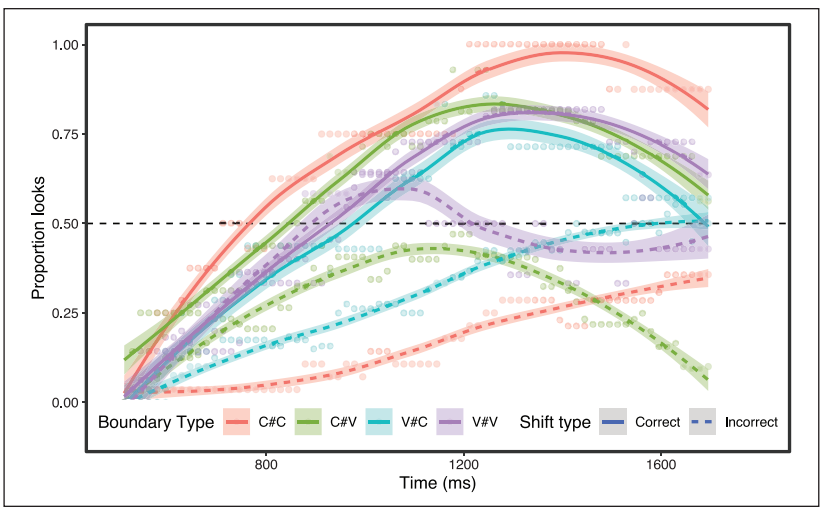

Figure 5. Onset-contingent growth curves for looks at target picture in the Test trials. Solid lines indicate correct shifts from distractor to target, and dashed lines indicate incorrect shifts from target to distractor.

V#V respectively. The dotted lines represent incorrect shifts for the four conditions and show a very high rate of incorrect shifts for the V#V boundary type, followed by less high rates in the C#V and V#C boundary types, and the lowest shift rate for the C#C boundary type. To model this data, we ran a logistic mixed-effects model on the onset-contingent data predicting the proportion of switches from the first fixated image as a function of the two Carrier Phrase × Target Word

Table 5. Onset-contingent analyses.

	β	SD	CI [95%]
Intercept	0.49	0.2	0.1,0.88
Carrier Phrase: Vocoid	−0.68	0.28	−1.2,−0.13
Target Word: Vowel	−0.31	0.3	−0.9,0.3
First Image: Target	−1.69	0.24	−2.2,−1.2
Carrier Phrase: Vocoid × Target Word: Vowel	0.37	0.44	−0.5,1.2
Carrier Phrase: Vocoid × First Image: Target	1.15	0.32	0.5,1.8
Target Word: Vowel × First Image: Target	0.16	0.36	−0.5,0.9
Carrier Phrase: Vocoid × Target Word: Vowel × First Image: Target	−0.44	0.55	−1.5,0.6

Note. M4: *Shift ~ Carrier Phrase × Target Word × First Image + (1 + Carrier Phrase × Target Word × First Image | Child) + (1 | Item)*. Values in the table are on the log-odds scale.

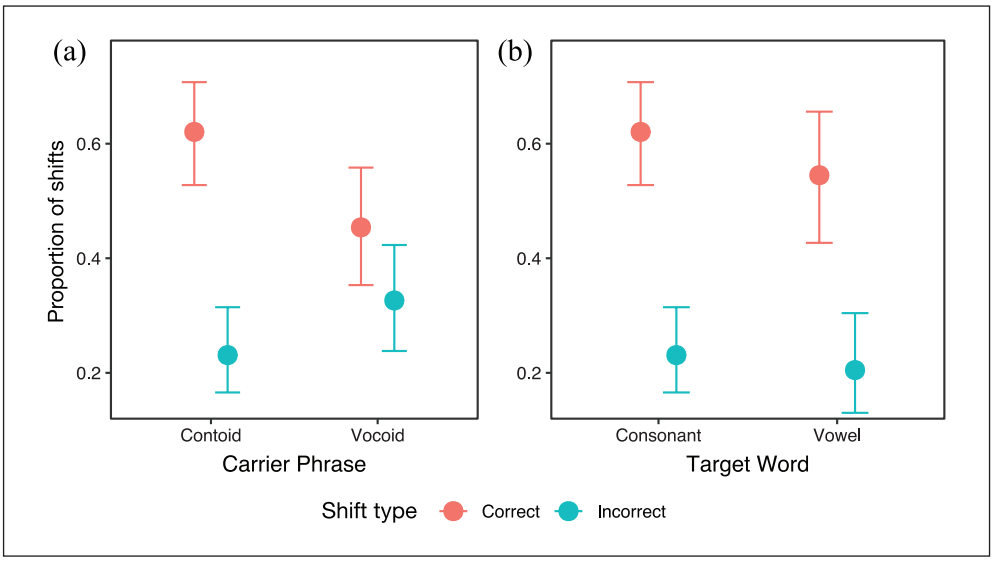

Figure 6. Marginal effects for the interaction between (a) Carrier Phrase and First Image and (b) Target Word and First Image. Values on the *y*-axis are in the scale of the response variable (proportions). Error bars represent 95% credible intervals.

boundary types, with the addition of a term for whether the first fixated image was the target image (i.e., incorrect shift) or the distractor image (i.e., correct shift), M4: *Shift ~ Carrier Phrase × Target Word × First Image + (1 + Carrier Phrase × Target Word × First Image | Child) + (1 | Item)*. Results from the model are reported in Table 5, and predicted values are plotted in Figure 6. The model fit showed a main effect of Carrier Phrase, suggesting that vocoid-final carrier phrases lead to a larger number of both correct and incorrect gaze shifts compared to contoid-final carrier phrases (see Figure 5), which was supported by decisive evidence ($\beta = -0.68$, SD = 0.28, CI[95%] = −1.2,−0.1, BF[CarrierPhraseVocoid < 0] > 100, Post.Prob. = 1), as well as a main effect of Target Word, suggesting that vowel-initial target words also lead to a generally larger number of gaze shifts than consonant-initial target words, although evidence in favor of this effect was less

strong (β = −0.31, SD = 0.3, CI[95%] = −0.9,0.3, BF[TargetWordVowel < 0] = 5.93, Post.Prob. = .86). More interestingly, the model showed an interaction effect between Carrier Phrase and First Image, revealing that vocoid-final carrier phrases lead to more incorrect gaze shifts and fewer correct shifts than is the case for contoid-final carrier phrases (β = 1.15, SD = 0.32, CI[95%] = 0.5,1.8, BF[CarrierPhraseVocoid:FirstAOILookAtTarget > 0] > 100, Post.Prob. = 1; as shown in Figure 6(a)), but no interaction between Target Word and First Image (β = 0.16, SD = 0.36, CI[95%] = −40.5,0.9, BF[TargetWordVowel:FirstAOILookAtTarget > 0] = 1.95, Post.Prob. = .66; as shown in Figure 6(b)). No credible three-way interaction between Carrier Phrase, Target Word, and First Image was found (β = −0.44, SD = 0.55, CI[95%] = −1.5,0.6, BF[TargetWordVowel:FirstAOILookAtTarget < 0] = 3.78, Post.Prob. = .79). We then ran the model again after having collapsed the first two predictors in the previous model (Carrier Phrase × Target Word) into one factor, Boundary Type, with four levels C#C, C#V, V#C, and V#V, in order to pair the rate of incorrect shifts more closely to the four boundary types, model structure: *Shift ~ Boundary Type × First Image + (1 + Boundary Type × First Image | Child) + (1 | Item)*. The model fit revealed that the C#C boundary was associated with the lowest rate of incorrect shifts (β = 0.53, SD = 0.21, CI[95%] = 0.13,0.95), followed closely by the C#V boundary (β = 0.28, SD = 0.37, CI[95%] = −0.46,1), which was followed in turn by V#C (β = 1.28, SD = 0.35, CI[95%] = 0.61,1.99) and V#V boundaries (β = 1.2, SD = 0.44, CI[95%] = −0.34,2.04).

4 Discussion

A number of studies showing that Danish-learning children lag behind in early linguistic development (Bleses et al., 2008; Bleses, Basbøll, Lum, et al., 2011; Bleses, Basbøll, & Vach, 2011) have raised the question as to whether the phonetic idiosyncrasies of Danish might have a direct negative impact on real-time speech processing, and therefore on learning. Spoken Danish is characterized by an opaque speech signal in which morphological and lexical boundaries often fall within strings of vocoids, which, compared to contoids, may constitute less reliable processing cues in speech (e.g., Mattys & Jusczyk, 2001; Nespor et al., 2003). We were therefore interested in testing the hypothesis that the recognition of familiar words in child-directed carrier phrases would be hindered by the presence of vocoids around the word boundaries in a sample of Danish-learning 24-month-olds.

In summary, our results showed that the presence of contoids at word boundaries resulted in more robust target word recognition, whereas vocoids had a detrimental effect on processing. This effect was driven by either the final segment of the carrier phrase or the initial segment of the target word being a vocoid. When looking at how proportional looks to target unfolded through the time window of a trial, we found particularly strong evidence that vocoid-final carrier phrases (i.e., V#C and V#V) resulted in a lower degree of proportional looking at target than contoid-final carrier phrases (i.e., C#C and C#V). Plotting correct shifts (i.e., distractor to target) against incorrect shifts (i.e., target to distractor) provided additional evidence that vocoid-final carrier phrases led to a higher number of incorrect shifts and to a lower number of correct shifts than contoid-final carrier phrases (i.e., the probability of making a correct shift was higher for C#C and C#V than for V#C and V#V). Moreover, we found strong evidence for an interaction effect of Target Word and Cubic Time, suggesting that children oriented less consistently and more slowly to vowel-initial target words (i.e., C#V and V#V) than to consonant-initial target words (i.e., C#C and V#C). These findings suggest that the presence of vocoids on either side of the target word boundary impeded real-time language processing in our task, although with a differential effect on processing depending on the position of the vocoid, with a stronger effect of carrier phrase-final vocoids (as discussed further below). Lastly, we found that target word recognition was especially hindered in

the presence of vocoids on both sides of the target word boundary: a bootstrapped cluster-based permutation analysis revealed a substantial 351-ms delay in orienting to the target image in V#V sentences compared with C#V sentences. Together, these findings also seem to suggest that the number of vocoids at word boundary is inversely proportional to target word processability, so that C#C is easier to process than C#V, which is easier than V#C, which is ultimately easier than V#V.

We propose an explanation of these findings in the light of incremental theories of language processing, such as the *Chunk-and-Pass* model (Chater & Christiansen, 2018; Christiansen & Chater, 2016). In this framework, language comprehension is to a large extent dependent on the listener's ability to quickly process the incoming speech signal by chunking it into coherent units and passing it on to increasingly higher levels of representations for further processing (e.g., from acoustic input to syllables, words, multiword chunks, phrases, and so on up to discourse-level representations). Because of the constraints put on online language processing by memory limitations on the one hand, and by the transience of spoken language on the other, this process of recoding is intrinsically "now-or-never": it must happen very rapidly, before the input is lost. Whereas the burden of dealing with such *Now-or-Never bottleneck* (Christiansen & Chater, 2016) in speech processing has generally been laid on the listener, we suggest here that phonetic characteristics of the language may also affect the chunk-and-pass process by making the speech signal inherently more or less "chunkable." In particular, we argue that contoids and vocoids may affect speech chunkability differently by virtue of their perceptual and distributional properties. The presence of consonants/contoids at word boundaries should promote the processing of the unfolding speech stream into coherent chunks (e.g., words) by making their boundaries more explicit. From a perceptuo-acoustic perspective, contoids provide robust speech segmentation cues by creating salient spectral discontinuities in an otherwise continuous and virtually steady-state signal (e.g., Lieberman et al., 1957; Stevens, 1998), while from a distributional perspective they seem to carry most of the weight of lexical information and thus facilitate the chunking process (e.g., Bonatti, Peña, Nespor, & Mehler, 2005; Nespor et al., 2003), at least in certain languages. By contrast, the presence of vocoids may impede the chunking process by making word boundaries less explicit, thus affecting the chunk-and-pass process. This should be particularly true of sequences of two or more adjacent vocoids, in which information about boundaries between units are less perceptually salient than alternating sequences of contoids and vocoids (e.g., Wright, 2004). Sequences of two or more vocoids lack the sonority cues that generally result from the alternation of low-sonority sounds (e.g., stop consonants) and high-sonority sounds (full vowels), resulting in the lack of sonority cues to syllable structure in Danish (e.g., Trecca et al., 2019; see also Basbøll, 2012).

Therefore, we further suggest that this adverse effect of vocoids on processing may be cumulative, so that longer sequences of vocoids (which are common in Danish) will generate a *bottleneck-driven delay* that accumulates as the sentence unfolds, ultimately slowing down processing. This may explain the fact that our results showed a larger detrimental effect of vocoids in carrier phrase-final position than in target word-initial position, given that the carrier phrase in question ("Her er. . .") was in fact a sequence of four adjacent vocoids. Following this explanation, the cumulative delay generated by the challenging chunking-and-passing of the adjacent vocoids might have taxed the processing system, ultimately impeding the recognition of the target word.

The hypothesized bottleneck-driven delay in processing target words when vocoids are present at the word boundary seems to prevail independently of the vowel-processing bias found for Danish-learning children when learning new words (Højen & Nazzi, 2016). The advantage for processing *paradigmatic* vowel distinctions in a CVC context (such as [dɛt] vs. [dæt]) found by Højen and Nazzi does not seem to transfer to an advantage for *syntagmatic* processing of successive vocalic sounds as cues for segmentation/chunking in fluent speech. That is, the bottleneck-driven delay in Danish may be due to an intrinsic, language-independent property of

vocoids in fluent speech that is not ameliorated by practice with vocalic contrasts in isolated words. Nonetheless, Danish-learning children may still be relying on their early exposure to a vocoid-rich ambient language, thus perhaps performing better on the task presented here than children learning more contoid-heavy language would (e.g., English, Italian, or French). This would suggest that Danish-learning children may be compensating to a certain extent, but not enough to perform as well on vocoid-laden sequences as on contoid-laden sequences in the present task. However, this conclusion is speculative and calls for additional (cross-linguistic) research specifically targeting the question of which features of vowels in Danish may be most detrimental for online speech processing.

Finally, we also suggest that the processing delay observed in this study may partly explain the delay observed in Danish children's acquisition of lexicon and morphology (e.g., Bleses et al., 2008; Bleses, Basbøll, & Vach, 2011). Marchman and Fernald (2008) found that the speed at which 25-month-old children oriented to target images when these were named in a LWL experiment predicted linguistic skills (expressive vocabulary and sentence complexity) as late as at 8 years of age. In their study, this predictive relation was mediated by the children's working memory skills (cf. MacDonald & Christiansen, 2002), suggesting that the ability to retain the incoming speech signal and process it promptly before the information is lost relates fundamentally to language learning (cf. Chater & Christiansen, 2018). In another recent LWL study of Danish-learning children, Trecca et al. (2018) showed that consonant-initial nonsense labels (e.g., *syffen*) that were presented to children in a vocoid-final carrier phrase (*Her er*. . .) were responded to less accurately in a later test phase, compared to when the same nonsense words were presented in a contoid-final carrier phrase (*Find*. . .). This finding integrates well with the results of the present study and offers appealing evidence in support of the idea that processing skills are contiguous with language learning.

Despite the encouraging results, there are nonetheless two methodological aspects of the present study that may limit some of the generalizability of our results, and thus seem worthy of addressing in future studies. One potential issue is a possible animacy effect related to the fact that both consonant-initial target words depicted inanimate objects (*car* and *teddy bear*), whereas both vowel-initial target words depicted animals (*monkey* and *duck*). This may have led to a spurious preference for vowel-initial targets (see e.g., LoBue, Bloom Pickard, Sherman, Axford, & DeLoache, 2013). Furthermore, all of the six words used in Filler trials were consonant-initial (three of which were /b/-initial), which may have biased processing in favor of the two consonant-initial target words (*bil* and *bamse*). Both animacy bias and filler-word bias resulted from our need to choose target words that (a) satisfied specific criteria in terms of their phonetic structure, (b) were comparable in terms of their frequency of occurrence in Danish child-directed speech, and (c) that were known by most Danish two-year-olds based on MB-CDI normative data for Danish (Bleses et al., 2008). A second potential issue relates to the lack of an explicit pre-naming phase in our trials. This is usually included in LWL studies in the form of a 2-second time window, in which the pair of pictures is presented to the child with no sound, before the speech stimulus is played. Together with the aforementioned animacy bias, the lack of a pre-naming phase may have contributed to producing spurious initial preferences for some of the target objects. Whereas the use of onset-contingent analysis helped us control for the effect of initial preferences in our study, any cross-linguistic comparison between our results and those of comparable studies should be done with caution.

5 Conclusion

Our results provide initial support for the hypothesis that the lack of consonants/contoids in relation to lexical boundaries in spoken Danish can impede speech processing in young

Danish-learning children (e.g., Bleses et al., 2008; Bleses & Basbøll, 2004). Characteristics of the language processing system on one hand (e.g., Chater & Christiansen, 2018; Christiansen & Chater, 2016), and the peculiar vocoid-heaviness and consequent opaqueness of the Danish sound structure on the other (e.g., Grønnum, 2003), may make processing of spoken Danish particularly challenging. Our findings seem to support this idea, despite previous evidence showing that Danish-learning children learn to use the large number of vowels to their advantage (Højen & Nazzi, 2016). If spoken Danish indeed is intrinsically harder to process than other comparable languages, Danish-learning children may need a higher degree of exposure to the same input before this can be appropriately processed and stored in memory, possibly explaining their delay in acquiring vocabulary and inflectional morphology.

While having the intelligibility of Danish as their point of departure, the implications of our findings may extend beyond the study of Danish and generalize cross-linguistically to mechanisms of language processing and acquisition. Whereas numerous linguistic and extralinguistic factors can be responsible for differences in rates of acquisition across languages, there is evidence suggesting that the ratio of contoids to vocoids in each language may explain part of this variation. Bleses, Basbøll, Lum, et al. (2011) found that the rate at which a number of European languages are acquired in the first two years of life is negatively correlated with the proportion of vocoids to contoids in their sound inventories: languages that resemble Danish in being particularly rich in vocalic sounds (such as Dutch and Swedish) were shown to be acquired at a slower rate than languages in which contoids outnumber vocoids (such as Croatian and English). The mainland Scandinavian languages (Danish, Norwegian, and Swedish) offer a particularly unique opportunity for testing questions related to sound structures and speech processing empirically, since the three countries provide for an almost-perfect natural experiment (Trecca et al., under review): the three languages are typologically and genetically very close and mutually intelligible, and the three countries of Denmark, Norway, and Sweden are very highly comparable in terms of societal and cultural factors. Future cross-linguistic studies may help isolate the unique effect of speech characteristics on language processing and learning, possibly confirming that features of speech that are more challenging for the language system to process may make some languages intrinsically harder to process and learn than others.

Acknowledgements

We thank Hans Basbøll for providing valuable feedback on our speech stimuli as well as on the interpretation of our results, and Sofie Neergaard for helping in the creation of the speech stimuli. We are also grateful for feedback from Teresa Cadierno, Charlotte Gooskens, and Padraic Monaghan, as well as from three anonymous reviewers on an earlier version of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by Danish Council for Independent Research (FKK) Grant DFF-7013-00074 awarded to MHC.

ORCID iDs

Fabio Trecca https://orcid.org/0000-0002-5900-7616

Anders Højen https://orcid.org/0000-0003-2923-5084

Supplemental material

Supplemental material for this article is available online and is accessible in the below URL: https://journals.sagepub.com/doi/suppl/10.1177/0023830919893390/suppl_file/SupplementalMaterialpdf.pdf

Notes

1. For a comprehensive description of the phonetics and phonology of Danish, we refer the reader to Grønnum (1998) and Basbøll (2005).
2. In order to keep the procedure as compact as possible, we did not have a pre-naming phase in which the two pictures are shown prior to the speech stimulus being played (as in, e.g., Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Hendrickson & Friend, 2013; Jesse & Johnson, 2016). Possible consequences of the lack of a pre-naming phase are considered in the Discussion.
3. CI[95%] = two-sided 95% Credible Intervals based on quantiles. The true β value falls within the CI in the posterior predictive distribution with 95% probability (see Bürkner, 2017).
4. Bayes Factor (BF) quantifies the ratio of the marginal likelihoods of two competing models or hypotheses. Following Kass and Raftery (1995), a BF of 1–3.2 is considered to not be worth more than a bare mention, a BF of 3.2–10 indicates substantial evidence, a BF of 10–100 indicates strong evidence, and a BF of above 100 indicates decisive evidence.

References

Basbøll, H. (2005). *The phonology of Danish*. Oxford, UK: Oxford University Press.

Basbøll, H. (2012). Monosyllables and prosody: The Sonority Syllable Model meets the word. In T. Stolz, N. Nau, & C. Stroh (Eds.), *Monosyllables: From phonology to typology* (pp. 13–41). Berlin, Germany: Akademie Verlag.

Benavides-Varela, S., Hochmann, J., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences, USA*, *109*, 17908–17913.

Bleses, D., & Basbøll, H. (2004). The Danish sound structure — Implications for language acquisition in normal and hearing impaired populations. In E. Schmidt, U. Mikkelsen, I. Post, J. B. Simonsen, & K. Fruensgaard (Eds.), *Brain, hearing and learning. 20th Danavox symposium* (pp. 165–190). Copenhagen, Denmark: Holmen Center Tryk.

Bleses, D., Basbøll, H., Lum, J., & Vach, W. (2011). Phonology and lexicon in a cross–linguistic perspective: the importance of phonetics—A commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language*, *38*, 61–68.

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past–tense studies. *Language and Cognitive Processes*, *26*, 1193–1231.

Bleses, D., & Trecca, F. (2016). Early acquisition of Danish in a cross-Scandinavian perspective: A psycholinguistic challenge? In H.-O. Enger, M. I. N. Knoph, K. E. Kristoffersen, & M. Lind (Eds.), *Helt fabelaktig! Festskrift til Hanne Gram Simonsen på 70-årsdagen* (pp. 13–28). Oslo, Norway: Novus forlag.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*, 619–650.

Bleses, D., Vach, W., Wehberg, S., Faber, K., & Madsen, T.O. (2007). *Tidlig kommunikativ udvikling: Værktøj til beskrivelse af sprogtilegnelse* [Early communicative development: A tool for assessing language development]. Odense, Denmark: Syddansk Universitetsforlag.

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*, 451–459.

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own name recognition at 5 months. *Developmental Science*, *18*, 587–598.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Chater, N., & Christiansen, M.H. (2018). Language acquisition as skill acquisition. *Current Opinion in Behavioural Sciences*, *21*, 205–208.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Science*, *39*, e62.

Dale, P., & Goodman, J. (2005). Commonality and individual differences in vocabulary growth. In M. Tomasello & D. I. Slobin (Eds.), *Beyond nature–nurture. Essays in honor of Elizabeth Bates* (pp. 41–80). London, UK: Lawrence Erlbaum.

Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazzi, T. (2014). Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1–15.

Dink, J. W., & Ferguson, B. F. (2015). eyetrackingR. R package version 0.1.8. Retrieved from http://www.eyetracking-R.com

Dussias, P., Valdés Kroff, J., Guzzardo Tamargo, R., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically-developing and late-talking toddlers. *Child Development*, *83*, 203–222.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*, 234–248.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology*, *60*, 190–217.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics. On-line methods in children's language processing* (pp. 97–136). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins Publishing Company.

Gooskens, C., van Heuven, V. J., van Bezooijen, R., & Pacilly, J. J. A. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication*, *52*, 1022–1037.

Grønnum, N. (1998). Danish: Illustrations of the IPA. *Journal of the International Phonetic Association*, *28*, 99–105.

Grønnum, N. (2003). Why are the Danes so hard to understand? In H. G. Jacobsen, D. Bleses, T. O. Madsen, & P. Thomsen (Eds.), *Take Danish – For instance* (pp. 119–130). Odense, Denmark: University Press of Southern Denmark.

Grønnum, N. (2005). *Fonetik og fonologi* [Phonetics and phonology]. Copenhagen, Denmark: Akademisk Forlag.

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, *14*, 439–456.

Hendrickson, K., & Friend, M. (2013). Quantifying the relationship between infants' haptic and visual response to word-object pairings. In *Proceedings of the 37th Annual Boston University Conference on Language Development, Boston University Conference on Language Development* (Vol. 37, Suppl). NIH Public Access.

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science*, *14*, 1445–1458.

Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science*, *19*, 41–49.

Jesse, A., & Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *Journal of Experimental Child Psychology*, *145*, 1–10.

Jørgensen, R. N., Dale, P. S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language*, *37*, 419–428.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 7732-795.

Kim, Y. J., & Sundara, M. (2015). Segmentation of vowel-initial words is facilitated by function words. *Journal of Child Language*, *42*, 709–733.

Kyhl, H. B., Jensen, T. K., Barington, T., Buhl, S., Norberg, L. A., Jørgensen, J. S., . . . Husby, S. (2015). The Odense Child Cohort: Aims, design, and cohort profile. *Paediatric and Perinatal Epidemiology*, *29*, 250–258.

Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

LoBue, V., Bloom Pickard, M., Sherman, K., Axford, C., & DeLoache, J. S. (2013). Young children's interest in live animals. *British Journal of Developmental Psychology*, *31*, 57–69.

MacDonald, M. C., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*, F9–F16.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.

Mattys, S., & Jusczyk, P. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception & Performance*, *27*, 644–655.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475–494.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*, 13–30.

Nazzi, T., Dilley, L., Jusczyk, A., Shattuck-Hufnagel, S., & Jusczyk, P. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language & Speech*, *48*, 279–298.

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*, 291–296.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 221–247.

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, *19*, 1223–1227.

Nishibayashi, L.-L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188–203.

Poltrock, S., & Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *131*, 135–148.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Schüppert, A., Hilton, N. H., & Gooskens, C. (2016). Why is Danish so difficult to understand for fellow Scandinavians? *Speech Communication*, *79*, 47–60.

Seidl, A., & Johnson, E. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, *35*, 1–24.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stokes, S. F., Bleses, D., Basbøll, H., & Lambertsen, C. (2012). Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech, Language, and Hearing Research*, *55*, 1265–1273.

Trecca, F., Bleses, D., Madsen, T. O., & Christiansen, M. H. (2018). Does sound structure affect word learning? An eye-tracking study of Danish learning toddlers. *Journal of Experimental Child Psychology*, *167*, 180–203.

Trecca, F., McCauley, S. M., Andersen, S. R., Bleses, D., Basbøll, H., Højen, A., . . . Christiansen, M. H. (2019). Segmentation of highly vocalic speech via statistical learning: Initial results from Danish, Norwegian, and English. *Language Learning*, *69*, 143–176.

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (under review). The puzzle of Danish: Implications for language learning and use.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. M. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). New York, NY: Cambridge University Press.