

Cue-weighting and (individual) pathways to the acquisition of vowel height contrasts: The F0 and F1 routes

Jérémy Genette*, University of Antwerp, Department of Linguistics, CLiPS Computational Linguistics and Psycholinguistics, Antwerp, Belgium, jeremy.genette@uantwerpen.be

Steven Gillis, University of Antwerp, Department of Linguistics, CLiPS Computational Linguistics and Psycholinguistics, Antwerp, Belgium

Jo Verhoeven, University of Antwerp, Department of Linguistics, CLiPS Computational Linguistics and Psycholinguistics, Antwerp, Belgium; City St George's, University of London, School of Health and Medical Sciences, London, United Kingdom

*Corresponding author.

This paper explores cue-weighting in the production of vowel height contrasts in children's first words. Specifically, it investigates the role of the first formant (F1) and the fundamental frequency (F0) in signalling these contrasts and how their use develops alongside lexical development. Over 12,000 high and low vowels were acoustically analysed for F0 and F1 in the spontaneous speech of 30 Dutch-speaking children. Particular attention is paid to the developmental trajectories of both cues and to the between-children variability in cue-weighting. The results provide a detailed picture of developmental trends in the production of primary and secondary acoustic cues in spontaneous early speech. Overall, the findings indicate that children gradually use primary (F1) and secondary (F0) cues more actively as their vocabulary expands. The results also suggest that children exhibit some degree of control over the production of the secondary F0 cue. Finally, the analyses reveal large inter-individual variability, with children following different developmental trajectories in their acquisition of phonetic secondary cues.



1. Introduction

The vowel systems of the world's languages are organised according to distinct phonological dimensions, each playing a crucial role in conveying linguistically relevant information. These dimensions include vowel height, place of articulation, lip rounding, tenseness, nasalisation, Advanced Tongue Root, and duration (e.g., Ladefoged & Maddieson, 1996; Laver, 1994). These parameters are typically described in articulatory, acoustic, or auditory terms (e.g., Chomsky & Halle, 1968; Jakobson et al., 1952; Ladefoged & Maddieson, 1996; Laver, 1994; Stevens, 1989). For instance, vowel height is traditionally described articulatorily by the degree of stricture inside the oral cavity (e.g., Laver, 1994; Rietveld & Heuven, 2009). Acoustically, vowel height is characterised by the frequency of the first formant (F1) (e.g., Delattre, 1951; Delattre et al., 1952; Lindau, 1975, 1978): high vowels such as [i] and [u] typically have a low F1 while [a] or [ɑ] have a high F1. Linking linguistically contrastive speech sound categories to their main acoustic correlate is characteristic of work at the interface between phonetics and phonology. Researchers identify phonological categories, such as the low vowels [a] and [ɑ], and the high vowels [i] and [u], and seek to characterise them in terms of discrete acoustic measurements, such as F1.

Although the number of linguistically relevant categories is limited, the speech signal carries a vast amount of acoustic information (cf. Puggaard-Rode, 2022). While F1 is the main correlate of vowel height, there is additional variation in the acoustic spectrum associated with changes in vowel height that potentially enhance the information conveyed by F1: different types of phonetic information often co-occur and work together in “bundles” to signal a single linguistic category (Stevens et al., 1986).

Efficiently distinguishing phonemic categories thus relies on the complex relationship between acoustic cues (Schertz & Clare, 2020). The contrast between speech sounds is not determined by a one-to-one correspondence between speech sound categories and phonetic components. This phenomenon is known as cue-weighting and is crucial for distinguishing linguistically contrastive units. While primary cues may be sufficient to determine category membership, it is more realistic to expect that multiple cues contribute in varying degrees to the identification of a given category (Lisker, 1978, 1986). For instance, in distinguishing vowel heights, the primary cue is typically F1, which is strongly predictive of vowel category. However, vowels of different heights also vary systematically in F0, which serves as a secondary cue to this contrast. The secondary cues can be (i) a physiological by-product of speech production mechanisms (Schertz & Clare, 2020, p. 8), or (ii) they might be actively controlled by speakers to ensure that the perception of the primary cue is enhanced (e.g., Kingston & Diehl, 1994, 1995; Kingston et al., 2008).

1.1. Cue-weighting in language acquisition

The acquisition of phonological categories by children is thus inherently related to the acquisition of multiple cues. Children must attend to and make use of multiple cues that convey phonological

contrasts. That is, children must learn to identify which of these cues are more reliable than others for signalling a specific contrast in their native language. Depending on their specific L1, children apply different strategies, giving more or less weight to phonetic features according to the linguistic relevance in their first language. For instance, Buder and Stoel-Gammon (2002) demonstrated that, in Swedish-speaking children, the acquisition of the linguistic relevance of vowel length leads to a progressively reduced reliance on final consonant voicing as a cue to vowel identity, giving more weight to vowel length. In contrast, American children do not show this effect as vowel length is linguistically less relevant for English-speaking children than for Swedish-speaking children.

In other words, to truly understand whether and how children attend to primary and secondary cues, it is now well-established that the development of cue weighting must be examined longitudinally, especially as development continues up to around 7 years of age (see Nittrouer, 1995).

Several studies have investigated how children's reliance on different cues in perception evolves over time. For example, Nittrouer (1992) examined the relative contribution of both formant transition cues and consonant-internal cues for English fricative contrasts in perception. She compared adults (20–40 years old) with groups of children aged 3, 5, and 7 years. The results showed that children are more sensitive to time-varying than to static cues and that their perceptual strategies develop from relying on whole-syllable forms to relying more on segment-independent information. In a later study, Nittrouer (2002) tested children aged 4, 6, and 8 years, as well as adults. The findings revealed that while children and adults assign similar weights to formant transitions for the /f/-/θ/ contrast, children present a different cue-weighting strategy for the /s/-/ʃ/ contrast: initially they prefer formant transitions but gradually shift towards the cues most relevant in their native language. Similarly, Walley and Carrell (1983) compared how adults and five year olds' weight onset spectra and formant transitions for different plosives in CV sequences. Their results indicated that both children and adults rely primarily on formant-transition information.

In speech production, McGowan et al. (2004) investigated the development of cues for [l] in children aged 14 to 26 months. The children achieved in non pre-vocalic positions adult-like productions only by the end of the time frame studied, indicating that control of the tongue needs time to mature. Hazan and Barrett (2000) studied older children aged 6 to 12 to determine at what age adult-like performance in phonemic categorisation emerges across a range of contrasts. Their results showed that adult-like weighting had not yet been fully achieved until between the ages of 10 and 22 years.

In conclusion, it has been shown that the importance, or weight, that is given to different cues varies with age (e.g., Nittrouer et al., 2013). However, very few studies have provided a detailed longitudinal account of how primary and secondary cues simultaneously evolve over time in early speech production. Moreover, certain aspects of cue-weighting that have been

documented in adult speech remain unexplored in child speech, such as the correlation between cues and the variability across individuals. After determining whether children use primary and secondary cues, the remaining gaps that this paper will address are the developmental pattern of each specific cue in early speech production, the correlation between the cues and the between-children variability.

1.1.1. The acquisition trajectory of secondary cues

There is a vast literature on how cue use and category boundaries change over time, but the focus has mostly been on perception, where several explanations have been proposed (see Cristia, 2008). In perception, the shift from children's to adults' weighting strategies has been called the developmental weighting shift (Nittrouer et al., 1993). According to Nittrouer (2002), this shift reflects children's initial focus on the general movements of the vocal tract, which helps them learn to reproduce these gestures before they acquire adult-like cue-weighting strategies. An alternative explanation (Sussman, 2001) attributes this shift instead to the maturation of the auditory system: children initially rely on cues that are perceptually more salient due to their length, amplitude, or contextual prominence. A third account argues that the shift is linked to the phonemic awareness acquired through reading instruction (Mayo & Turk, 2005).

However, the acquisition of sound contrasts also encompasses production, and in that respect, little is known to date. The nature of the link between perception and production is still much debated, and the results of studies comparing both modalities are complex (see Schertz & Clare, 2020, for a review). This is true for adults, but it remains largely unstudied in children, with Idemaru and Holt (2013) as a notable exception: they showed an increased use of the secondary cue in production, but no specific link between modalities.

Leaving aside the relation between modalities in the acquisition of sound contrasts, different pathways can explain how children develop cue-weighting in speech production. The acquisition of primary cues requires gaining control over specific articulatory gestures, while the acquisition of secondary cues is less straightforward and depends on the origin and nature of those secondary cues. Physiologically determined secondary cues differ from actively controlled ones: the usage of the former is directly correlated with the primary cue. In contrast, when secondary cues are actively controlled to enhance contrasts, their association with primary cues may emerge more gradually, reflecting the progressive acquisition and implementation of phonological categories by individual children. In addition to controlling the production of different cues, children must also develop adult-like sensitivity to their relative weighting in perception in order to use them with adult-like weightings in their own production.

To cut a long story short, more work on production is needed to obtain a clearer view of the time course of acquisition of primary and secondary cues (see also Samuel & Kraljic, 2009). This is precisely what the present paper seeks to address through a longitudinal, fine-grained study of spontaneous speech production.

1.1.2. Relationship between cues

Another crucial aspect of cue-weighting is that multiple cues may covary to jointly signal specific linguistic categories. While cue-weighting often results in a positive or negative correlation across categories, this does not necessarily imply that the same correlation holds within categories. Three possibilities present themselves (see Clayards, 2017, for a review).

First, two cues may share the same articulatory or physio-acoustic origin. For instance, increasing supraglottal volume by lowering the larynx is used to produce voicing in CV sequences, but it is also reported to result in a lower F0 onset on the following vowel (Hombert et al., 1979; Hoole & Honda, 2011; Hoole et al., 2004). As such, it influences both VOT and the F0 onset of the following vowel. In such cases, the correlation between the primary cue (VOT) and the secondary cue (F0 onset) is expected to correlate consistently with each other, i.e., the longer the negative VOT, the lower the F0 onset. Crucially, this same correlation should be observed within categories: different tokens of the same phoneme should show the same relationship in the same direction.

Second, multiple cues do not necessarily result from the same articulatory gesture and can be under the active control of the speaker (Kingston & Diehl, 1994). In this scenario, cues may be used to enhance the perception of a category. Returning to the preceding example, speakers might actively exaggerate F0 onset values to reinforce the perception of voicing that is already cued by VOT. This goes beyond the intrinsic relationship between physiologically linked cues. Here, a correlation between cues must exist across categories (e.g., voiced vs. unvoiced), but no consistent correlation is expected within categories.

Third, cues may also be used to actively compensate for the ambiguity of another cue when signalling a specific category. For example, if the primary cue (e.g., VOT) is ambiguous with respect to voicing, a secondary cue (e.g., F0) may be overtly enhanced to reinforce the target category and compensate for the ambiguous information provided by the primary cue (see also Repp, 1982). In this case, the within-category and the across-category correlation between cues are in opposite directions.

Determining whether such mechanisms are at play in children provides crucial insights into how secondary cues are integrated into their phonological knowledge and it has important implications for language acquisition. Whether children exert active control over secondary cues is key to understanding the developmental trajectory of cue use in speech production. This is precisely what the present paper aims to address.

1.1.3. Between-individual variability

In adult speech, substantial variability in cue use for a given contrast has been well documented not only across languages (Lisker & Abramson, 1964) and dialects (Schertz et al., 2019) but also within homogeneous groups of speakers. For instance, Kapnoula et al. (2017) demonstrated

considerable between-speaker variability in how categorical the categorisation of stimuli based on VOT and F0 is for each speaker. Similarly, Schertz et al. (2015)'s study of fortis-lenis distinctions has shown marked variability in how speakers rely on cues, particularly in perception. Along the same lines, Clayards (2017) reported significant between-talker variation in the relative use of VOT and F0 to signal voicing contrasts.

While between-speaker variability has been relatively well described in adults, much less is known about the degree of variability in children's speech production, which is inherently characterized by high variability (e.g., McGowan et al., 2004). Documenting this variability is crucial for understanding the degrees of freedom during the acquisition of primary and secondary cues, and for clarifying to what extent children differ in their cue-weighting strategies for realising a given contrast. Do all children consistently rely more on the primary cue, or do some children rely more on the secondary cue? And do these cues follow similar developmental trajectories across children? These are the questions that need to be addressed in order to better understand the phonetic realisation of phonological categories in early speech production. This paper aims to explore these questions by investigating Intrinsic Vowel Pitch.

1.2. Test case: Intrinsic Vowel Pitch

To investigate these three aspects, the present study focuses on the acquisition of the vowel height contrast, which is conveyed by the first formant (F1) as a primary cue and the fundamental frequency (F0) as a secondary cue. It is well established that the main acoustic correlate of vowel height is F1: [+high] vowels are more strongly associated with lower F1 values than [-high] vowels (e.g., Delattre et al., 1952; Miller, 1953). However, other cues, such as F0, may also contribute to signalling vowel height: high vowels tend to have a higher F0 than low vowels. This phenomenon of Intrinsic Vowel Pitch is robust and has been universally attested (see Whalen & Levitt, 1995, but see Connell, 2002). While Intrinsic Vowel Pitch has long been considered a purely biomechanical consequence of vowel articulation (e.g., Honda, 1983; Rossi & Autesserre, 1981), research suggests that Intrinsic Vowel Pitch may well undergo phonologization in some languages and become a linguistically relevant cue for vowel height contrasts (e.g., Diehl, 1991; Diehl & Kluender, 1989; Kingston, 1991).

The relevance of F0 to the perception of vowel height is demonstrated in several studies. Syrdal and Gopal (1986) investigated the acoustic cues of vowel height using modifications of natural stimuli and found that the most relevant acoustic dimension for the perception of vowel height is the distance between F1 and F0. In small F1-F0 distances, the auditory system averages the two spectral peaks which are located close to one another (cf. Chistovich & Lublinskaya, 1979) and the resulting low F1 value causes the vowel to be perceived as high. Conversely, when the distance is large, both peaks are perceived as distinct, resulting in the perception of the vowel

as low. The importance of the F1-F0 distance was further illustrated by Traunmüller (1981), Fahey and Diehl (1996), and Fahey et al. (1996).

In short, F0 consistently contributes to signalling vowel height contrasts alongside F1. The presence and development of the F0 cue in early speech production have been examined in three studies: Bauer (1988), Whalen et al. (1995), and Gregory et al. (2013). However, all three focus on prelexical vocalisations and, as such, do not capture how children progressively learn to associate these cues with phonological categories in lexical productions.

1.3. The present work

The overarching goal of this paper is to investigate whether children have already developed F0 as a cue to vowel height as they start to produce their first words and to document its development up to the age of two, alongside the development of F1 as the primary cue. To this end, the study investigates the developmental trajectory of F1 and F0 cue-weighting in the early speech of 30 Dutch-speaking children, using a large-scale longitudinal study of spontaneous speech samples. The research questions addressed in this paper are as follows:

1. Do children use F1 as a primary cue only to distinguish vowel height in their first lexical words or do they also use F0 as a secondary cue?
2. What is the developmental trajectory of the F1 and F0 cues to vowel height and are they independent of each other?
3. How do the F1 and F0 cues correlate at the token and speaker levels?
4. What is the between-infant variability in the use of F1 and F0? That is, do children differ in the way they rely on the primary and secondary cues?

2. Methods

2.1. Vowel system of Standard Belgian Dutch

Belgian Standard Dutch is spoken by approximately six million speakers (Verhoeven, 2005). The vowel system of Belgian Standard Dutch consists of 12 monophthongs [i, ɪ, e:, ε, a, a:, ɔ, o:, u, ʏ, y:, ø:] and three diphthongs [ej, œy, ɔu]. The main difference between Belgian Standard Dutch and Netherlandic Dutch pertains to the pronunciation of the half-close long vowels [e:, o:, ø:], which are realised as closing diphthongs in the Netherlands (Gussenhoven, 1992, 2007; Verhoeven, 2005). Traditionally, four levels of vowel height describe the Dutch vowel system: close [i, u, y:], half-close [ɪ, e:, o:, ʏ, ø:], half-open [ε, ɔ], and open [a, a:]. Booij (1995)'s phonological description of the Dutch vowel system uses the features [\pm high] and [\pm mid] to distinguish the close [+high, -mid], half-close [+high, +mid], half-open [-high, +mid], and open [-high, -mid] levels.

2.2. Participants and recording procedure

The speech corpus used in the present study is the CLiPS Child Language Corpus (CCLC), which consists of manually transcribed longitudinal recordings of Dutch-acquiring infants, between the ages of 6 and 24 months, during spontaneous interactions with their caretakers (Molemans, 2011; Van den Berg, 2012; Van Severen, 2011). A total of 30 children were included as participants in this study. All children participated in all 19 recording sessions, except for three, who each missed one session.

During each session, the children interacted with their parents in their familiar home environment, engaging in daily routines or free play as they normally would. All parents were native speakers of Dutch and came from a mid-to-high socioeconomic status (SES) background. The recordings lasted 64 minutes on average. From those recordings, 20-minute samples were selected in which the children were most vocally active for analysis (Molemans, 2011, p. 31). The children were all hearing-typical and had no indication of health or developmental problems, according to parental reports and regular observations by the Flemish agency *Child and Family (Kind & Gezin)*. The administration of the N-CDI tests (Zink & Lejaegere, 2002) at 12, 18 and 24 months of age confirmed the normal language development of the children.

Except for vegetative and (dis-)comfort sounds, each utterance was identified and labelled. The procedure outlined by Vihman and McCune (1994) was applied to classify the children's utterances as prelexical (i.e., canonical babble or other vocalisations) or lexical (i.e., words).

2.3. Transcription and vowel identifications

The recordings were transcribed following the CHAT conventions using CHILDES' CLAN (MacWhinney, 2000). Subsequently, the transcriptions were converted into the TextGrid format suitable for Praat (Boersma & Weenink, 2021). Trained phoneticians transcribed the children's productions both with a broad phonetic transcription and phonemically, based on lexical identity. The analysis presented in this paper is based on the phonetic transcriptions, as the children's actual production can diverge considerably from the target lexical items. For instance, some children might consistently produce the lexical item /bal/, "ball" in Dutch, as [bɪl], due to the way it is encoded in their phonological system. In such cases, it would not be accurate to associate the acoustic F1 and F0 values with the /a/ category. For this reason, all the analyses are based on the broad phonetic transcription. The children's utterances that overlapped with any other speaker's utterance by more than 5 ms were excluded. The centre of the vowels was chosen as a measurement point (see Genette et al., 2023). They were detected by means of an adapted version of the syllablenuclei_v3 Praat script (de Jong et al., 2021). The Praat script identifies syllabic nuclei within the audio file of an utterance, relying on various parameters. Using the Parselmouth API (Jadoul et al., 2018) of Praat (Boersma & Weenink, 2021), a Python script tuned the hyperparameter `Minimum_dip_near_peak_(dB)` of de Jong et al. (2021)'s script via a binary search algorithm to meet the number of vowels expected

based on the manual transcription of the utterance. Some of the children's utterances could not be analysed automatically. In such cases, the vowels in those utterances were located manually. The total number of vocalic nuclei in the corpus amounted to 138,623 vowels.

2.4. Acoustic analysis

Initially, the F0 and F1 were extracted using Praat's standard parameters at the designated centres of vowels: Time step (s) = 0.0, Pitch floor (Hz) = 75 Hz, Very accurate = off, Pitch ceiling (Hz) = 600 Hz, Silence threshold = 0.03, Voicing threshold = 0.45, Octave cost = 0.01 per octave, Octave-jump cost = 0.35, Voiced / unvoiced cost = 0.14 for F0 and for F1, Time step (s) = 0.0, Maximum number of formants = 5, Formant ceiling (Hz) = 5500, Window length (s) = 0.025, Pre-emphasis from (Hz) = 50.

In a second iteration, the F0 was recalculated by adjusting the pitch floor and pitch ceiling for each child and each session using the formulae from Keelan et al. (2011), as in Equation (1) and Equation (2), based on the measurements obtained with the standard parameters. The formant ceiling was also increased to 8,000 Hz, as suggested in Praat's user manual for young children.

$$(1) \quad F0_{\text{ceiling}} = q65_{F0_standard_parameters} \times 1.9 + 10$$

$$(2) \quad F0_{\text{floor}} = q35_{F0_standard_parameters} \times 0.72 - 10$$

The F0 and F1 measurements were carried out in Hz, but a z-score normalisation (Lobanov, 1971) was used to reduce the potential effects of anatomical differences between children. Additionally, the measurements were also normalised within each recording session to account for anatomical differences due to the maturation of the different vocal structures of the children between age 6 and 24 months. To avoid biases resulting from the missing and unbalanced data that inevitably characterise naturalistic speech corpora, Lobanov (1971)'s normalisation was adapted via a logistic regression approach (Barreda & Nearey, 2018). As such, F1 and F0 were normalised separately by the regression modelling in R (R Development Core Team, 2022) presented in Equation 3, in which G represents the frequency in Hz (F1 or F0), S corresponds to the unique combination of child and session, and V represents the vowel category.

$$(3) \quad \text{model} = \text{lm}(G \sim 0 + S*V, \text{contrasts} = \text{list}(V = \text{contr.sum}))$$

The speakers' estimates and standard errors were extracted from the model. The intercept served as the per-subject and per-session mean in the Lobanov (1971) normalisation, respectively. In order to get per-subject and per-session standard deviations for the Lobanov (1971) normalisation, the respective standard errors determined by the model were multiplied by the square root of the number of observations, i.e., the number of vowels for that speaker and session.

2.5. Exclusion of outliers and data selection

Vowels for which no acoustic measurements could be obtained due to limitations of the tracking procedure were excluded from the analysis. In total, 20,457 vocalic tokens had to be excluded, which accounted for approximately 15% of all items. To exclude potential outliers, vowels characterised by acoustic measurements which were higher or lower than 2.5 standard deviations of the per-speaker and per-session average for that measurement were also excluded (cf. Garellek & Esposito, 2023). On this basis, a total of 6,340 such tokens were excluded from the analysis. In total, roughly 20% of the items were excluded from the analysis.

From the remaining items, all lexical utterances were extracted, yielding a total of 28,663 vowels. The average number of vowels produced by the children was 955.43 (SD = 463.58). From those lexical utterances the high vowels, [i] and [u], and the low vowels, [a] and [ɑ], were extracted. The final selection consisted of 12,906 vowels, and the total counts per vowel category and per child are presented in **Figure 1**. On average, each child produced the following: 46.833 [i] tokens (SD = 40.118), 64.867 [u] tokens (SD = 50.49), 205.133 [a] tokens (SD = 89.84), and 113.367 [ɑ] tokens (SD = 67.72).

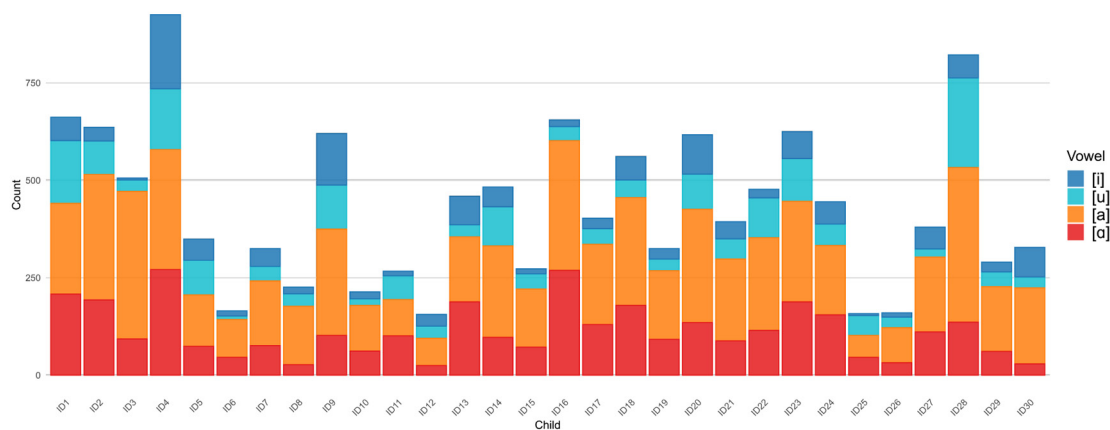


Figure 1: Total counts of vowels produced per child, split by vowel categories.

2.6. Language acquisition measure

To determine whether the use of F1 and F0 changed as lexical development progressed, each child's cumulative vocabulary was computed for each recording session. Expressive cumulative vocabulary is a standard estimate of children's increase in vocabulary in the spontaneous longitudinal recordings (e.g., Rowe et al., 2012; Vanormelingen et al., 2016). It is computed by counting the appearance of each new word type iteratively for each recording session. This measure reflects the observed first usage of lexical items during sessions rather than their absolute first usage.

The CLAN software (MacWhinney, 2000) was used to compute children's expressive cumulative vocabulary based on the transcripts. The first use of each lemma per child was marked with a Python script. Lemmas were added cumulatively to each child's vocabulary list recording session by session. At 11 months, the children had an average cumulative vocabulary of 1.22 (SD = 0.63), while at two years old, they had an average cumulative vocabulary of 220.14 (SD = 87.57). This indicates that, on average, children's usage of at least one lexical item was documented at 11 months, while by the last recording session, the children were observed to use an average of 220 lemmas. The cumulative vocabulary values were transformed to a natural logarithmic scale to prevent outliers from dominating the analyses (cf. Balling & Baayen, 2012).

2.7. Statistical analysis

The statistical analysis was carried out in R (R Development Core Team, 2022) with the R package *lme4* (Bates, 2015). The *lmerTest* package (Kuznetsova et al., 2017) was used to obtain p-values. Binomial mixed effect models of increasing complexity were built incrementally, including fixed and random effects, with vowel height coded as "high" vs. "low" as the dependent variable. Vowels were classified as low if transcribed as [a] or [ɑ], and as high if transcribed as [i] or [u], based on the transcriber's judgment of phonetic accuracy. To establish whether the inclusion of an effect significantly improved the fit between the predicted and observed values, a likelihood ratio test was carried out (see Baayen, 2008; Bates, 2015). The fixed effects that were tested included the main effects of normalised F1, normalised F0, logarithmic cumulative vocabulary, and its quadratic effect, as well as vowel place of articulation (with [a] and [i] categorised as front vowels, and [ɑ] and [u] as back vowels). Place of articulation was included because differences in lingual articulation may affect laryngeal tension, and previous studies have shown that it can influence the use of F0 to distinguish vowel height (Van Hoof and Verhoeven, 2011; but see Verhoeven and Connell, 2024). Potential interactions among these variables were also examined. The tested random effects included random intercepts and slopes per child. The final model is described in the results section. More details about the model construction can be found in the supplementary materials.

3. Results

In this study, the relative contribution of F1 and F0 to the vowel height contrasts was examined in the first words of children up to the age of two. **Table 1** presents the descriptive statistics of F1 and F0 for high and low vowels, both in Hz and on the normalised scale. The density distributions of F1 and F0 on the normalised scale for each vowel height category are also illustrated in **Figure 2**. As expected, the F1 of high vowels was lower than that of low vowels in Hz and on the normalised scale. Conversely, the F0 of high vowels was higher than that of low vowels in Hz and on the normalised scale, although this difference was smaller than that observed for F1. These results suggest that both cues correlate with vowel height in the expected direction.

| | Vowel Category | Frequency in Hz | | Normalised Frequency | |
|----|----------------|-----------------|-----------|----------------------|-----------|
| | | Mean | Std. Dev. | Mean | Std. Dev. |
| F1 | High | 636.29 | 213.87 | -0.05 | 0.18 |
| | Low | 1026.71 | 292.11 | 0.23 | 0.24 |
| F0 | High | 413.07 | 84.25 | 0.11 | 0.22 |
| | Low | 372.38 | 96.84 | 0.04 | 0.21 |

Table 1: Descriptive statistics of F1 and F0 on the Hz and normalised scale for high and low vowels.

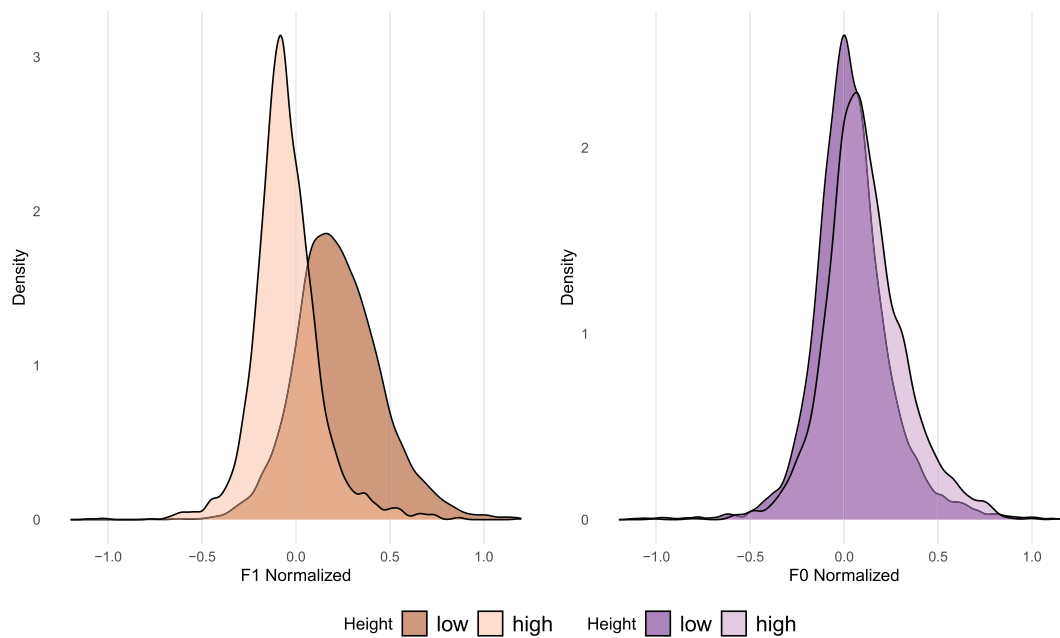


Figure 2: Density distribution function of F1 (left) and F0 (right) for high and low vowels.

A main objective of the present study is also to examine the development of the use of the F1 and F0 cues to signal vowel height contrasts. **Figure 3** shows the observed normalized F1 (left) and normalized F0 (right) values for each investigated vowel as a function of vocabulary growth. Visual inspection of the graphs reveals that the two high vowels (light and dark blue) gradually diverge from the two low vowels (orange and red). This pattern is observed for both F1 and F0. However, the contrast appears to emerge more rapidly and dramatically in F1 than in F0.

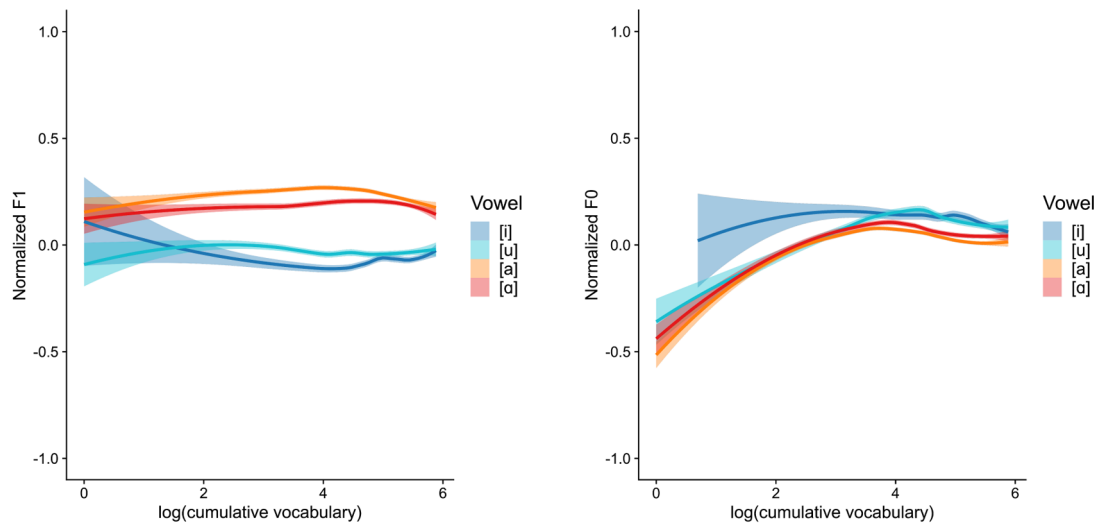


Figure 3: Observed normalized F1 (left) and normalized F0 (right) value per vowel category during lexical development as measured by cumulative vocabulary.

These findings are supported by the results of the binomial logistic regression. The results of the fixed effects from the modelling procedure are presented in **Table 2**. The final model includes fixed main effects of F1, F0, cumulative vocabulary, and place of articulation (“front” for [a] and [i] vs. “back” for [ɑ] and [u]). Additionally, it comprises the interaction between F1 and F0 as an interaction effect alongside the main effects of F1 and F0. The model also includes interactions between F1 and cumulative vocabulary, between F0 and cumulative vocabulary, between both cues and place of articulation, as well as a three-way interaction between F1, F0, and place of articulation. Furthermore, the model incorporates random intercepts per child and random slopes for the effects of F1, F0, cumulative vocabulary, and place of articulation. This structure allows the effects of these predictors to vary across individual children. As a sanity check, variance inflation factors (VIFs) of the predictors were calculated. There was only mild collinearity between some predictors. None had to be excluded following the guidelines of Montgomery and Peck (1992), since all VIFs were below five, and most were close to one.

In the remainder of this section, the marginal effects of the predictors of interest for the different research questions are discussed one by one. Typically, binary dependent variables, as in the present study, lead to nonlinearities in the predicted probabilities, and marginal effects are useful to summarize the effect of an independent variable as they can summarize multiple related coefficients, reduce the scaling issues inherent to logit-based models and can easily be expressed on another scale than the model coefficients (see Long & Freese, 2001; Mize, 2019).

| Predictor | Estimate | SE | z value | p value |
|---------------------------------|----------|-------|---------|---------|
| (Intercept) | -2.009 | 0.323 | -6.215 | <.001* |
| F1 | -1.877 | 0.703 | -2.668 | 0.008* |
| F0 | 1.666 | 0.594 | 2.806 | 0.005* |
| F1 : F0 | -1.870 | 0.622 | -3.005 | 0.003* |
| Log(cumulative vocabulary) : F1 | -1.426 | 0.149 | -9.595 | <.001* |
| Log(cumulative vocabulary) : F0 | 0.405 | 0.116 | 3.488 | <.001* |
| F1 : Place of articulation | 0.795 | 0.175 | 4.555 | <.001* |
| F0 : Place of articulation | -0.949 | 0.147 | -6.468 | <.001* |
| Log(cumulative vocabulary) | 0.273 | 0.071 | 3.859 | <.001* |
| Place of articulation | 0.524 | 0.079 | 6.651 | <.001* |
| F1 : F0 : Place of articulation | 1.166 | 0.589 | 1.978 | 0.048* |

Table 2: Results of the fixed effects for the best-fitting model.

The first research question aims to investigate whether F0 is a predictor of vowel height, similar to the established correlation between F1 and vowel height at the beginning of lexical production by children. The observed effects of F0 ($\beta = 1.666$, $SE = 0.594$, $p = 0.005$) and of F1 ($\beta = -1.877$, $SE = 0.703$, $p = 0.008$) are significant and in the expected direction. To better understand the effects of F0 and F1, marginal effects plots are presented in **Figure 4** showing the model's output transformed from logits to percentages using the inverse logistic function, with standard deviations represented by shaded areas. The left panel of **Figure 4** illustrates the curvilinear and categorical effect of F1 on the probability that a vowel is classified as high. As expected, lower F1 values are associated with a higher probability of being categorised as high, while lower F0 values are associated with a lower probability of being categorised as high. In other words, children use a high F0 alongside a low F1 to signal a high vowel in their first words before the age of two. Note that those effects are observed at $\text{Log}(\text{cumulative vocabulary}) = 0$, indicating that both cues are already significant cues to the vowel height contrasts in the children's first words.

Beyond these main effects, the modelling procedure reveals significant interactions between both acoustic cues and place of articulation ("back vs "front"); this is illustrated in **Figure 4** by means of solid and dotted lines. Significant interactions are observed between F1 and place of articulation ($\beta = 0.795$, $SE = 0.175$, $p < .001$) and between F0 and place of articulation ($\beta = -0.949$, $SE = 0.147$, $p < .001$). The left panel shows that front and back vowels present relatively similar probability curves. However, the curve for front vowels appears slightly more categorical: i.e., the change from a high vowel (with a predicted probability close to 100%) to a

low vowel (with a predicted probability close to 0%) occurs over a narrower range of F1 values. In other words, a smaller change along the F1 dimension is sufficient to signal the phonological contrast for front vowels, whereas achieving the same contrast for back vowels requires a larger phonetic contrast. A comparable interaction effect is observed for F0. The probability curve for front vowels is again more categorical than that for back vowels. This indicates that the span of F0 values required to change a low vowel (with a predicted probability close to 0%) into a high vowel (with a predicted probability close to 100%) is shorter for front vowels than for back ones. Consequently, back vowels require a greater degree of phonetic differentiation in terms of F0 to convey the same phonological contrast, which is in line with previous findings (e.g., Van Hoof & Verhoeven, 2011). In summary, the interactions suggest that the effects may differ slightly for front and back vowels, while the direction of the effect remains consistent. Importantly, place of articulation does not interact with vocabulary size in the present data: although place of articulation affects the magnitude of the F0 effect, developmental changes are not analysed separately by place of articulation.

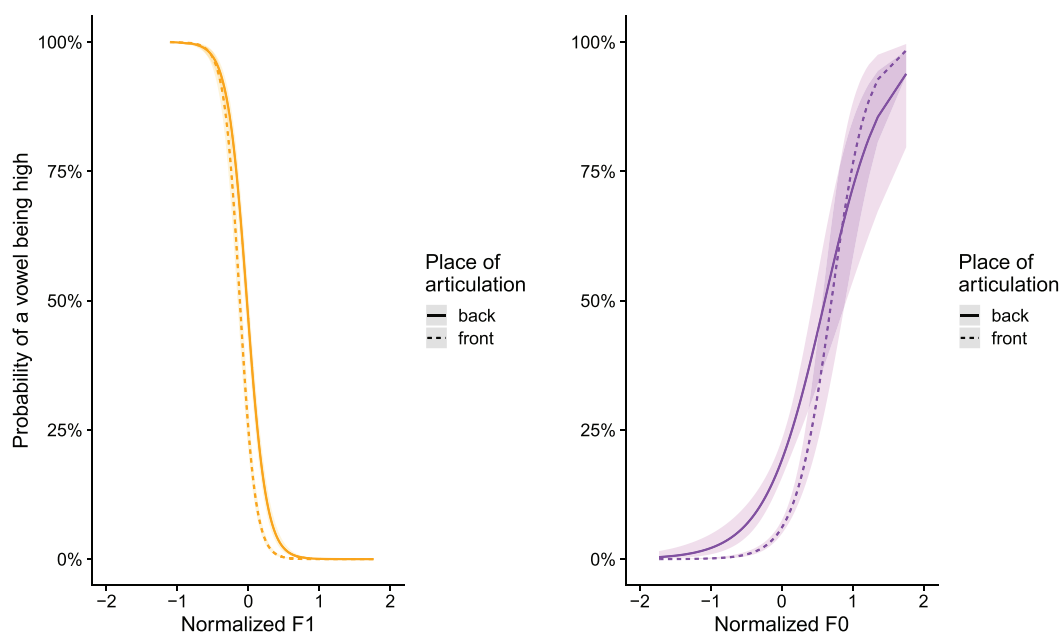


Figure 4: Marginal effect plot showing the predicted probability that a vowel is high as a function of F1 (left) and F0 (right) for front (dotted lines) and back (solid lines) vowels, as well as their standard deviations (shaded areas).

The second research question focuses on developmental change, specifically on the developmental trajectories of F1 and F0 as cues to vowel height. These trajectories can be examined by analysing how the marginal effects of F1 and F0 vary as a function of cumulative

vocabulary growth. **Figure 5** illustrates this developmental trajectory with a so-called spotlight analysis (Spiller et al., 2013). The left panel shows the effect of cumulative vocabulary on the probability that a vowel will be high for different levels of F1 (low, average, high). The same information is provided for F0 in the right panel.

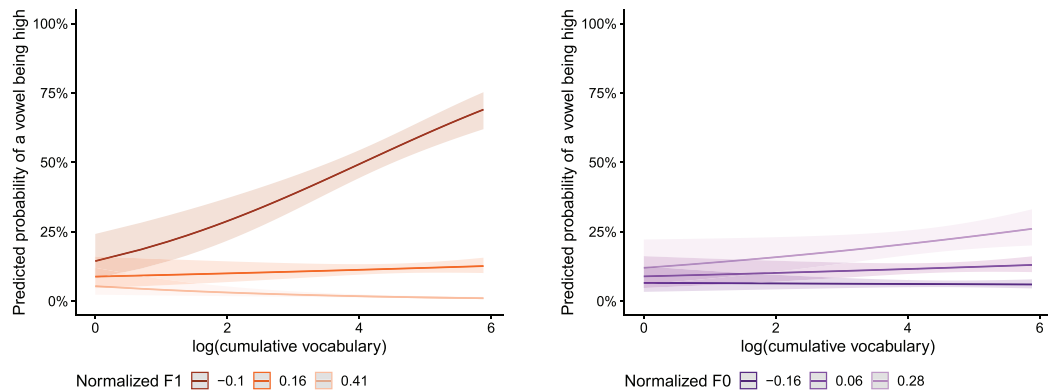


Figure 5: Visualisation of predicted probability that a vowel is high as a function of F1 (left) and F0 (right), and their standard deviations (shaded areas) during lexical development as measured by cumulative vocabulary.

As to F1, even in the earliest stages of lexical development (i.e., children's use of their first words), a vowel with a low F1 (i.e., -0.1 on the normalised scale) has a 14% probability of being categorised as a high vowel. As children's vocabulary expands, this probability further increases to 69%. Conversely, a very high F1 (i.e., 0.41 on the normalised scale) is initially associated with a low probability (5%) of being a high vowel, which further decreases to only 1% as lexical development progresses. This suggests that the association between low F1 and high vowels becomes stronger as children acquire more words, indicating that F1 is a reliable cue for vowel height early on but that its importance increases over time.

The developmental trajectory of F0 is less pronounced than that of F1. The probability that a vowel with a high F0 (i.e., 0.28 on the normalised scale) is a high vowel increases from 12% at the onset of lexical development to 26%. However, the probability that a vowel with a low F0 (i.e., -0.16 on the normalised scale) is a high vowel goes down from 7% to 6% throughout development: while F0 contributes to vowel height categorisation, its influence develops more slowly and is less consistent compared to F1.

To determine more precisely at what stage of lexical development children begin to distinguish high and low vowels based on the F1 and F0 cues, the Johnson-Neyman technique (Johnson & Neyman, 1936) is used to analyse the marginal effects of the interaction between

cumulative vocabulary and those two cues. The results are plotted in **Figure 6**. It shows the slope effect of the normalised F1 (left panel) and F0 (right panel) and their confidence intervals as a function of the logarithm of cumulative vocabulary: that is, it shows whether lexical development influences the way the two cues predict vowel height. The more the slope departs from 0 (see the horizontal dotted line), the larger the cumulative vocabulary has an effect on how F1 and F0 signals the probability that the vowel is high. In other words, the more the slope departs from 0, the larger the developmental change. If the confidence intervals include 0, it means that there is no significant change in the effect of lexical development on the probability that the vowel is high based on F0 or F1. On the contrary, if the confidence intervals do not include 0, it means that the lexical development changes the way F0 or F1 predicts vowel height.

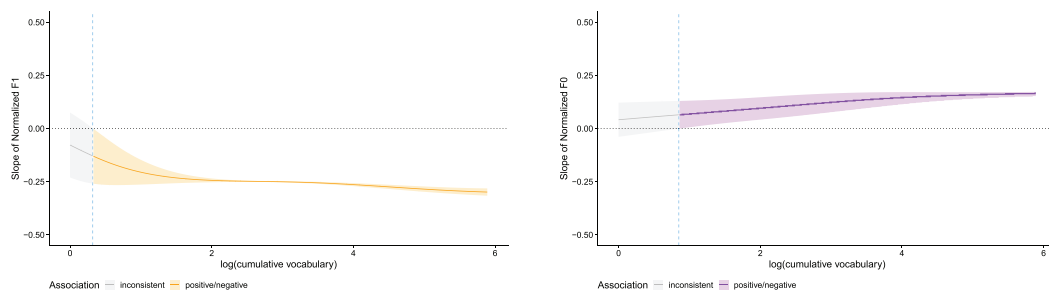


Figure 6: The marginal effects of the interaction between cumulative vocabulary and F1 (left) or F0 (right) on the probability of producing a high vowel and the lower and upper bounds of the confidence bands (shaded areas).

As such, the left panel of **Figure 6** shows that cumulative vocabulary significantly affects the way F1 influences the probability of a high vowel. It shows that the children's lexical development has an effect on the association between the phonetic cue and the phonological category. More specifically, cumulative vocabulary had such an effect from 0.32 of logarithmic cumulative vocabulary onwards, i.e., when the children reached a cumulative vocabulary of 1.38. This means that as soon as infants begin using at least one word during any recording session (i.e., not necessarily their very first word), they start to increasingly distinguish between high and low vowels using the F1 cue. In contrast, the right panels show that the association between F0 and cumulative vocabulary only becomes significant when the children reach a cumulative vocabulary of two words (more specifically, 2.34 or 0.85 on the logarithmic scale). This implies that it takes a bit longer for children to begin to actively enhance the F0 cue to signal vowel height distinctions. Through development, children start using this cue more reliably to enhance the perceptual contrast between vowel categories, but the change occurs more progressively and gradually as depicted by the flatter curve for the F0 slope than for the F1 slope.

The third research question addresses the question of the correlation between the primary F1 and secondary F0 cues to vowel height. The interaction effect proves to be significant ($\beta = -1.87$, $SE = 0.622$, $p < 0.01$), meaning that the higher the value of F1, the more the effect of F0 is reduced. It should be noted that the present interaction might suggest that the relationship between the F0 cue and vowel height is not (entirely) a physiological by-product of vowel articulation, since no interaction is usually expected if this were the case. However, to better uncover how both cues correlate together, the observed data were analysed similarly to Clayards (2017) by looking at the correlation between both cues across and between categories for each token. Because high vowels have, on average, a lower F1 and a higher F0, a negative between-category correlation is expected. The same analysis was carried out at the level of the speaker means, and the results of both are reported in **Table 3**.

| Level of analysis | Data | Pearson R | p value | N |
|----------------------|------|-----------|---------|--------|
| <i>Token</i> | all | 0.092 | < .001* | 12,906 |
| | high | 0.098 | < .001* | 3,351 |
| | low | 0.217 | < .001* | 9,555 |
| <i>Children mean</i> | all | -0.493 | < .001* | 60 |
| | high | -0.030 | .874 | 30 |
| | low | 0.267 | .153 | 30 |

Table 3: Pearson *R* correlations for normalised F1 and normalised F0 across and within each vowel height category for both individual tokens and children means.

A weak but significant positive correlation is found across categories between F1 and F0 ($r[12904] = 0.092$, $p < .001$). Within categories, a weak but significant positive correlation is also exhibited by high vowels ($r[3349] = 0.098$, $p < .001$) and a significant positive correlation is observed for low vowels ($r[9553] = 0.217$, $p < .001$). This is consistent with what Clayards (2017) refers to as an intrinsic relationship between the two cues.

However, the positive correlation across categories is surprising and might indicate that individual factors obscure the internal structure of the categories. When looking at the children's average, a strong significant negative correlation is indeed observed ($r[60] = -0.493$, $p < .001$). Within vowel height categories, no significant correlation is found for high vowels ($r[30] = -0.030$, $p = .874$) and low vowels ($r[30] = 0.267$, $p = .153$). The presence of a between-category correlation and the absence of a within category correlation might indicate active control by the speakers who signal the vowel height contrast and is not likely to result from a simple biomechanical consequence of the primary articulatory gesture (see Clayards, 2017).

However, these insights are based on the observed data and do not account for the hierarchical or longitudinal nature of the data. The fourth research question addresses this by focusing on between-individual variability and individual learning trajectories as captured by the statistical model coefficients.

The findings of the previous research questions have shown that the use of F0 as a secondary cue is related to the individual use of the children, but to get further details, it was tested whether children who rely more heavily on F1 also make greater use of F0. However, the individual coefficients derived from the model do not indicate this. Rather, there appears to be no correlation between use of the primary and secondary cues at the individual level. This interpretation is supported by **Figure 7**, which displays the individual F1 and F0 coefficients (the final model was refitted with log-transformed cumulative vocabulary, centred at the group mean). The plot, together with the Pearson correlation coefficient ($r[28] = -0.001$, $p = .996$), shows no significant association between cues. This suggests that children who use F1 efficiently to cue vowel height do not necessarily show the same efficiency with F0, and vice versa. Such dissociation between cues further implies that they cannot be attributed to a single articulatory gesture, since a shared gesture would likely exert a similar influence on both primary and secondary cues for each child. This means that children already use an individual cue-weighting strategy to signal vowel-height contrasts before two years of age.

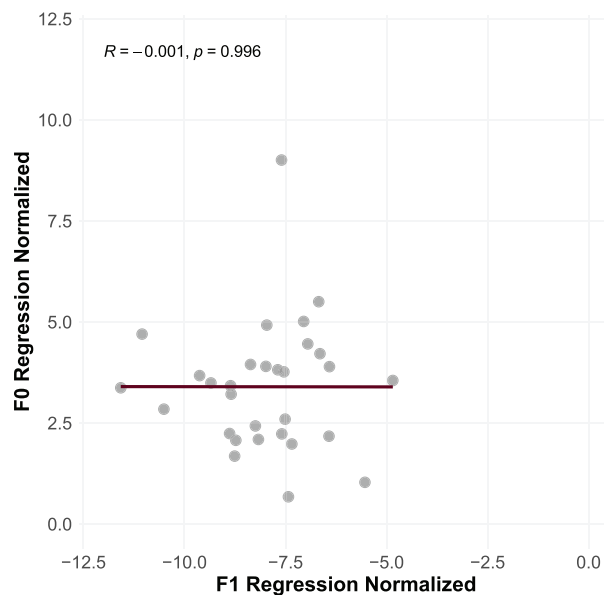


Figure 7: Scatterplot showing the individual F0 coefficients as a function of the corresponding F1 coefficients for each child in the corpus, with the fitted correlation line. Coefficient estimates are computed at the mean-centred value of Log(cumulative vocabulary).

For a more fine-grained understanding of developmental changes in cue-weighting, **Figure 8** demonstrates how each child in the corpus uses both cues at an earlier and a later time point in the recordings. To this end, the model was fitted with $\text{Log}(\text{cumulative vocabulary})$ centred at 0 and its maximum value. Besides showing that children increase their use of both cues from their earliest productions (empty circles) to the latest productions recorded in this study (full circles), the results reveal that the use of the secondary cue by a child is not necessarily correlated with the use of the primary cue. In **Figure 8**, individual coefficients are ordered by increasing F0. As such, the fact that the use of the F1 cue does not follow a similar trend further suggests that use of both cues is not correlated.

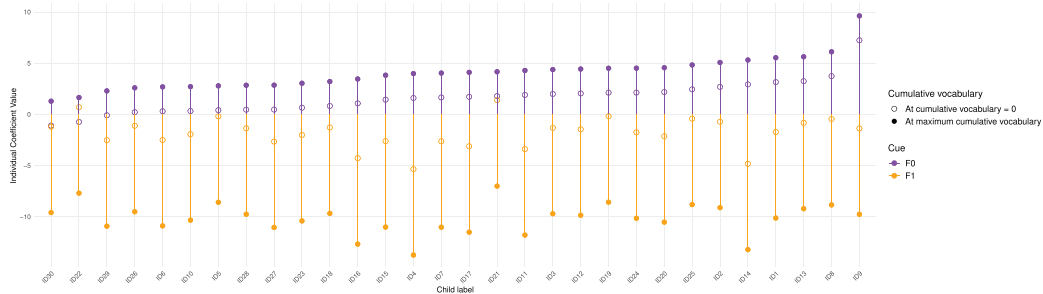


Figure 8: Lollipop chart showing between-children variability in the use of F1 (orange) and F0 (purple) as predictors of vowel height. Empty circles represent individual coefficients at cumulative vocabulary = 0, while full circles represent coefficients at maximum cumulative vocabulary.

The previous analysis shows how infants rely on F1 and F0 at the start and end of the recording sessions, but did not account for individual variability in developmental trajectories. To address this, two other analyses were carried out. First, we examined the correlation between the individual random slopes of F1 and F0, which was -0.02 , indicating that the development of the F1 cue was not correlated with the development of the secondary cue at the individual level. Second, to investigate the variability of the developing trajectories, the individual developmental trajectories of the F1 and F0 effects were examined. The individual coefficients for each cue per child were computed as a function of cumulative vocabulary. **Figure 9** illustrates these effects: the left panels display the probability of a vowel being high with a low F1 (set to mean F1 – SD) and a high F1 (set to mean F1 + SD) as a function of cumulative vocabulary, while F0 is held constant at the mean observed F0 value. Conversely, the right panels present the probability of a vowel being high with a high F0 (set to mean F0 + SD) and a low F0 (set to mean F0 – SD), while F1 is set at the mean observed F1 value in the corpus. The plots of those marginal effects serve to examine how variations in F1 and F0 affect the probability that a vowel is high as lexical development progresses. In other words, from the upper panels, it can be inferred whether a low F1 or a high F0 (the expected cues of a high vowel) are associated with a high probability (i.e.,

closer to 100%) that the vowel is high. Conversely, the lower panels show whether a high F1 or a low F0 (the expected cues of a low vowel) are associated with a low probability, closer to 0%, that the vowel is high. The closer the coefficients in the upper panels were to 100% the more a low F1 or high F0 correctly cue a high vowel. In the lower panels, the closer a high F1 and a low F0 were to 0% the more reliably they cue a vowel as low.

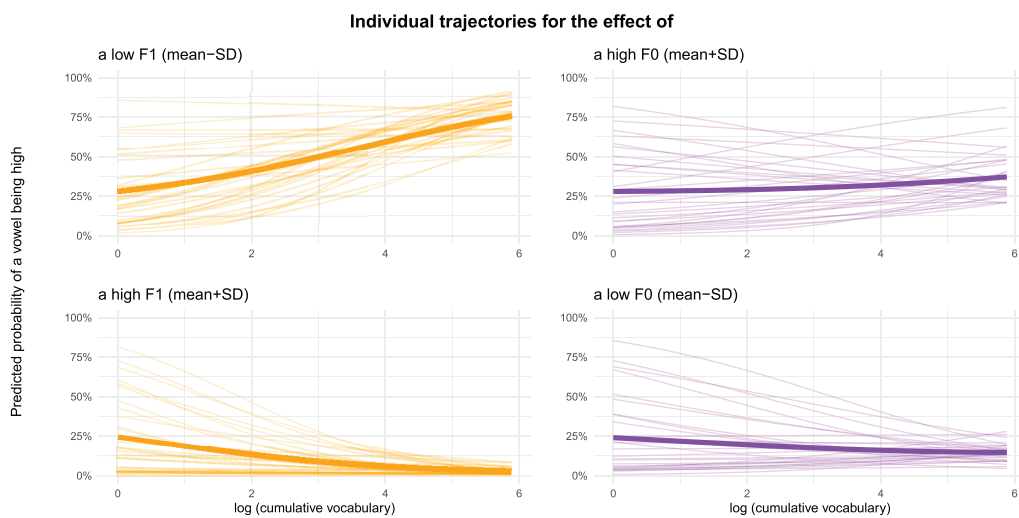


Figure 9: Individual coefficients for the F1 (left) and F0 (right) developmental effects. Each line corresponds to a child; the bold line indicates the group average.

A comparison of the variability in the use of the F1 cue (orange panels) versus the F0 cue (purple panels) reveals that children appear to converge more consistently in the use of F1. In the latest stage of lexical development, the probability that children used a high F1 to signal a low vowel was between 1% and 9%, whereas a low F1 indicated a high vowel with a higher probability, ranging from 53% to 91%. The convergence between individual developmental trajectories is particularly evident later in development, and greater variability among children was observed at earlier stages. A similar pattern can be observed for the F0 cue, though with greater variability than for F1, even at more advanced stages of lexical development. In the last observed stage of lexical development, the probability that a vowel is high with a high F0 ranged between 19% and 81%, and the probability that a vowel is low with a high F0 ranged between 4% and 32%. Similar to the F1 cue, the earlier the stage in lexical development, the higher the variability.

4. Discussion

This study investigates the relative use of F1 and F0 to distinguish high and low vowels in a group of 30 typically developing children acquiring Belgian Dutch. The main objective is to contribute to the current understanding of cue-weighting by providing a unique description of

cue use in the production of vowel height contrasts in children under the age of two, based on a time-dense, large-scale corpus of spontaneous speech. The children were recorded monthly in spontaneous interactions with their caretakers. A total of 12,906 high and low vowels were analysed acoustically for F0 and F1. This paper addresses four research questions: (i) Do children use F1 as a primary cue only to distinguish vowel height in their first lexical words or do they also use F0 as a secondary cue?; (ii) What is the developmental trajectory of the F1 and F0 cues to vowel height?; (iii) How do F1 and F0 correlate with each other?; and (iv) What is the between-infant variability in the use of F1 and F0?

The answer to the first research question is that the children in the study used both F0 and F1 to distinguish vowels of different phonological vowel heights in their first words. A high F0 and a low F1 cue a high vowel, while a low F0 and a high F1 cue a low vowel. This finding is consistent with many studies that indicate F1 is a strong cue to vowel height (e.g., Delattre, 1951; Miller, 1953). Moreover, this finding is consistent with the large body of research suggesting that F0 is also a cue to vowel height (e.g., Ewan & Krones, 1974; Fischer-Jørgensen, 1990; Honda, 1983; Whalen et al., 1995; but see Connell, 2002). The results show that as early as their first words, children use F0 to signal vowel height contrasts. This aligns well with the results obtained by Whalen et al. (1995) in babbling. These results indicate that F0, as a secondary cue to vowel height, is likely to be physiologically determined to some extent, but it cannot be excluded that children control it to some extent.

The modelling procedure further reveals interaction effects between place of articulation and both F1 and F0. Specifically, the high-low vowel contrast appears to be phonetically implemented through a larger difference in F1 in back vowels than in front vowels. This pattern is broadly consistent with studies reporting higher F1 values for /u/ than for /i/ in several languages, including Portuguese (Escudero et al., 2009), American English (Peterson & Barney, 1952; Strange et al., 2007), French (Strange et al., 2007) and Spanish (Cervera et al., 2001). However, opposite patterns have been reported in cross-linguistic work by de Boer (2009, 2011), and these studies do not explicitly compare the F1 distance between high and low vowels for front and back vowels. As such, the underlying sources of the observed effect of place of articulation on the F1 cue remain unclear and cannot be decisively established within the scope of the present study.

As to the observed interaction between F0 and place of articulation, the significance of the effect is consistent with previous findings. That is, the F0 contrast between high and low vowels is larger for back vowels than for front vowels. This pattern corroborates previous research (Van Hoof & Verhoeven, 2011), even if it is not always statistically significant (e.g., Verhoeven & Connell, 2024; Whalen & Levitt, 1995). The varying significance can arguably be attributed to differences in the data (and languages) analysed, as well as to differences in methodology. The effect itself, however, could be explained in several ways. On the one hand, it could be of articulatory origin, with lingual and jaw articulation differing between front and back vowels and

exerting forces on the larynx to different extents (see Chen et al., 2021). On the other hand, the difference in F0 between high and low back vowels might be larger than that observed for front vowels for perceptual reasons. If the need to distinguish vowels motivates F0 use, the presence of additional cues may influence whether F0 is needed to further aid vowel discrimination. As such, front vowels might benefit from the F2 cue, as they also differ along that dimension, whereas the F2 might help less in distinguishing back vowels.

The second research question examines the developmental trajectory of the F1 and F0 cues. Overall, high and low vowels gradually become more distinct from each other as the difference between the F1 and F0 of both vowel categories increases. However, the contrast in F0 appears to increase slightly later. That is, children distinguish vowels primarily by means of F1 (the primary cue) from early on but begin to distinguish vowels by F0 (the secondary cue) slightly later in their lexical development. These results have two implications. First, children progressively learn to use F0 as a way to increase the contrast between high and low vowels. Second, the developmental trend in the use of F0 does not align with that of F1, suggesting that the acquisition of the F0 contrast is, at least to some extent, the result of linguistic learning rather than a purely biomechanical phenomenon, as also demonstrated by the significant interaction between F1 and F0. This provides evidence that children younger than two years have already acquired secondary cues and are progressively developing them. It also suggests that Intrinsic Vowel Pitch might be at least partially perceptually motivated (Fahey & Diehl, 1996; Fahey et al., 1996; Traunmüller, 1981). However, this issue needs further investigation.

The third research question investigates the correlation between F1 and F0. The results show that there was, on a token-by-token basis, a weak but positive correlation across and within categories. However, the positive sign of the weak correlation is not expected, as higher F0 values are expected to occur with lower F1 values. This might be because speakers do not correlate F1 and F0 on a vowel-by-vowel basis but do so on an individual basis. That is, speakers might be unable to modulate cues for each token, but they might, on average, associate a high F1 value with a low F0 value and vice versa. However, speakers may vary in the extent to which they rely on each of the two cues. In fact, at the speaker level, a significant negative correlation is found across categories while no correlation was observed between categories. The expected negative correlation between F1 and F0 appears at the individual level and only across categories, suggesting that the Intrinsic Vowel Pitch might be under some active control of the speakers (e.g., Fahey & Diehl, 1996; Fahey et al., 1996; Traunmüller, 1981) rather than a mere biomechanical byproduct (e.g., Honda, 1983; Rossi & Autesserre, 1981). This was further demonstrated by the analysis of the individual coefficients of the F1 and F0 effects extracted from the model. Besides giving insights into the potential nature of Intrinsic Vowel Pitch, it demonstrates the early ability of children to actively control fine phonetic details and secondary cues to signal phonological contrasts.

The results of the fourth question aimed to document the between-children variability. The individual coefficients reported in **Figure 9** highlight the high degree of between-children variability and indicate that there is no direct correlation between individual F1 and F0 cues, which might indicate that both cues do not derive from the same articulatory gesture. It cannot be excluded, however, that the absence of correlation between the individual coefficients points to a great amount of individual variability. The present study cannot pinpoint the exact source of the variability. The higher degree of variability might also be the result of the other linguistic functions that F0 can serve in speech, such as for prosody. It also cannot be definitively established whether the observed between-children variability is due to differences in the extent to which children encode F0 as a secondary cue into their phonological system, or whether it stems from their phonetic implementation of the F0 cue.

A closer longitudinal inspection of the individual coefficients throughout lexical development reveals that the developmental trajectories for the acquisition of the F0 contrast were more varied than for conveying this same contrast with F1, especially for high vowels. One potential explanation is that the secondary nature of F0 leads to a later and less robust acquisition compared to the primary cue for vowel height. However, the source of this variability cannot be firmly established. This finding highlights promising avenues for future research aimed at decomposing the variance observed in the realisation of secondary cues in early speech.

In short, the observation of cue-weighting between F1 and F0 shows that infants could use multiple cues to produce a single contrast from a very young age and that they are also capable of precise laryngo-oral coordination. The results show how the correlation between F1 and F0 emerges at the individual level between categories but not within categories, which seems to indicate that even so early in development, children may control for this cue. Taken together, the results reveal how children can develop actively-controlled secondary cues to enhance phonemic contrasts early in their linguistic development. Moreover, the results provide insights suggesting that Intrinsic Vowel Pitch (i.e., the use of F0 as a secondary cue to vowel height) is likely to partially be of physiological origin, but it is also a cue that must be acquired and actively controlled. However, this requires further investigation.

4.1. Limitations and future work

Several studies have investigated F0 as a cue to vowel height contrasts in adults (e.g., Connell, 2002; Ewan & Krones, 1974; Fischer-Jørgensen, 1990; Honda, 1983; Whalen & Levitt, 1995), and just two older studies on children (Bauer, 1988; Whalen et al., 1995). This study is the first to evaluate the contribution of F1 and F0 cues simultaneously in the production of the vowel height contrast in young children (see Genette et al., 2025, for school-aged children) while also accounting for individual variability.

Nevertheless, some limitations of the present study must be pointed out. First, the naturalistic data collection inevitably resulted in some noise in the recordings. To minimise any influence of this on the measurements, careful data cleaning was carried out by fine-tuning the extraction parameter per speaker and per session and excluding outliers, which excluded a substantial part of the children's productions. Furthermore, the detection of the vowel centres was automated, most probably resulting in some detection errors. The applied procedure also did not enable dynamic analyses, as only the centre of the vowel was located.

It must also be acknowledged that measuring F1 in voices characterised by high pitch might result in some biases, as F1 might be overestimated for high vowels and under-estimated for lower vowels (Chen et al., 2019; Shadle et al., 2014).

In addition, the present study relied on perceptual transcriptions of vowels, based on the transcribers' judgments of phonetic accuracy. The results nevertheless show how the acoustic signal encodes what is perceived as distinct phonological categories and, as such, whether the acoustic cues of the children point towards a specific phonological category more or less efficiently. As such, the results indicate that, whether due to an increasingly clearer contrast between vowel categories, or a more consistent use of acoustic cues that might result from fewer off-target productions, the realisation of the contrast improves over time. Yet it is still unclear to what extent variations in the use of both cues are functionally relevant. That is, the present study cannot address whether vowel contrasts are significantly less intelligible in children who rely predominantly on F1 rather than both F1 and F0. This opens promising avenues for future research, such as investigating the extent to which mismatches between target phonemes and phonetic accuracy can be explained by differences in cue use. While the relationship between the intended target vowel and the expression of primary and secondary acoustic cues falls outside the scope of the present analysis, it opens promising avenues for future research.

It must also be emphasised that the results presented here are specific to Dutch and cannot be straightforwardly generalisable to other languages without further investigation, as the use of F0 as a cue to vowel height may differ across languages. Some languages may or may not take advantage of secondary cues to distinguish vowel height distinctions, as demonstrated by Van Hoof and Verhoeven (2011). Similar research should therefore aim to replicate the present study in other languages, especially those with less crowded vowel inventories than Dutch, since vowel inventory size is known to affect the relevance of the F0 cue for vowel height contrast. Furthermore, Esling et al. (2019) suggest that children learning languages in which the larynx carries a heavy functional load, such as pitch-dependent tone languages, might comparatively give more attention to precise laryngeal control. Comparisons with tone languages might thus be of interest for a better understanding of how language-specific laryngeal functional load affects the acquisition of F0 control for vowel height contrasts.

A limitation that should also be mentioned is that F0 serves functions beyond signalling vowel height contrasts, such as prosody, suprasegmental structure, or even emotional state, and this additional variability cannot be fully controlled. This variability might be levelled out by the size of the corpus in the present study, but further studies should investigate in more detail variability in F0 as a secondary cue to vowel height in early speech production.

Additionally, more research on this topic is needed to better understand how variability in the use of both cues becomes relevant later in the acquisition trajectory, as previous research has shown that the development of secondary cues can extend well into childhood, up to around seven years of age (e.g., Nittrouer, 1995). Further research should also look into the observed differences between front and back vowels in the use of F1 and F0 cues to distinguish high and low vowels.

5. Conclusion

This paper set out to investigate the production of phonemic contrasts by children under the age of two, specifically focusing on cue-weighting in the production of vowel height contrasts during the earliest stages of lexical development. Naturalistic data were collected on a monthly basis from 30 typically developing children. The vowels produced in lexical utterances were analysed acoustically for F0 and F1. The results of the analysis demonstrate that children consistently rely on F1 as a primary acoustic cue to distinguish vowels of different heights, but it is also clear that children can use F0 as a secondary cue for contrasting vowel heights from their first words onward.

The main contribution of the present study is the fine-grained documentation of developmental trends in the production of primary and secondary cues in spontaneous early speech. The vowel contrast between high and low vowels is found to gradually increase as children's vocabularies expand, as reflected by both the primary cue (F1) and the secondary cue (F0). In addition, the study describes how these cues interact and shows that the correlation between high F1 and low F0 values, or vice versa, emerges across categories, but not within categories, at the individual level. This finding suggests that even at an early stage of development, children have at least some degree of control over the production of secondary phonetic cues. Furthermore, the results present the high between-children variability in the use of the secondary cue. It is also shown that children vary in their reliance on the different cues: those who rely more heavily on the primary cue do not necessarily rely more on the secondary cue. This further describes high individual variability in the acquisition of several cues following different developmental trajectories.

In conclusion, this study demonstrates and describes cue-weighting in production for vowel height contrasts in the speech of children younger than two years, based on a time-dense, large-scale corpus of spontaneous speech.

Additional files

The additional files for this article can be found as follows:

- **File 1.** A. Model construction details. DOI: <https://doi.org/10.16995/labphon.17955.s1>
- **File 2.** An R Markdown document that was used to process the data and carry out the analysis and visualizations for this study. DOI: <https://doi.org/10.16995/labphon.17955.s2>

Acknowledgements

Thanks are due to the families and infants of this study. We would like also to thank K. Schauwers, I. Molemans, R. van den Berg and L. Van Severen for collecting the CLiPS Child Language Corpus. The research was supported by the Research Foundation in Flanders (FWO) [grant G004321N]. We are grateful to the reviewers for their valuable comments and suggestions that improved the quality of this work.

Competing interests

The authors declare that they have no competing interests.

Author contributions

Jérémy Genette contributed to conceptualization, methodology, formal analysis, and writing of the original and final drafts. Steven Gillis and Jo Verhoeven contributed to conceptualization, as well as to review and editing of the manuscript.

References

- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Balling, L., & Baayen, R. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, 125(1), 80–106. <https://doi.org/10.1016/j.cognition.2012.06.003>
- Barreda, S., & Nearey, T. (2018). A regression approach to vowel normalization for missing and unbalanced data. *Journal of the Acoustical Society of America*, 144(1), 500–520. <https://doi.org/10.1121/1.5047742>
- Bates, D. (2015). *Lme4: Mixed-effects modeling with R*. Springer.
- Bauer, H. (1988). Vowel intrinsic pitch in infants. *Folia Phoniatrica*. <https://doi.org/10.1159/000265901>
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer (Version 6[1.38]) [Computer software]. <http://www.praat.org>

- Booij, G. (1995). *The phonology of Dutch*. Oxford University Press.
- Buder, E., & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America*, 111(4), 1854–1864. <https://doi.org/10.1121/1.1463448>
- Cervera, T., Miralles, J., & González-Àlvarez, J. (2001). Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *Journal of Speech, Language, and Hearing Research*, 44(5), 988–996. [https://doi.org/10.1044/1092-4388\(2001/077\)](https://doi.org/10.1044/1092-4388(2001/077))
- Chen, W., Whalen, D., & Shadle, C. (2019). F0-induced formant measurement errors result in biased variabilities. *Journal of the Acoustical Society of America*, 145(5), 360–366. <https://doi.org/10.1121/1.5103195>
- Chen, W., Whalen, D., & Tiede, M. (2021). A dual mechanism for intrinsic f0. *Journal of Phonetics*, 87, 101063. <https://doi.org/10.1016/j.wocn.2021.101063>
- Chistovich, L., & Lublinskaya, V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, 1(3), 185–195. [https://doi.org/10.1016/0378-5955\(79\)90012-1](https://doi.org/10.1016/0378-5955(79)90012-1)
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.
- Clayards, M. (2017). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica*, 75(1), 1–23. <https://doi.org/10.1159/000448809>
- Connell, B. (2002). Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics*. <https://doi.org/10.1006/jpho.2001.0156>
- Cristia, A. (2008). Cue weighting at different ages. *Purdue Linguistics Association Working Papers*, 1, 87–105.
- de Boer, B. (2009). Why women speak better than men (and its significance for evolution). In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 255–265). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199545872.003.0014>
- de Boer, B. (2011). First formant difference for /i/ and /u/: A cross-linguistic study and an explanation. *Journal of Phonetics*, 39(1), 110–114. <https://doi.org/10.1016/j.wocn.2010.12.005>
- de Jong, N., Pacilly, J., & Heeren, W. (2021). Praat scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476. <https://doi.org/10.1080/0969594X.2021.1951162>
- Delattre, P. (1951). The physiological interpretation of sound spectrograms. *PMLA*, 66(5), 864–875. <https://doi.org/10.2307/459542>
- Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesised from spectrographic patterns. *WORD*, 8(3), 195–210. <https://doi.org/10.1080/00437956.1952.11659431>
- Diehl, R. (1991). The role of phonetics within the study of language. *Phonetica*, 48(2–4), 120–134. <https://doi.org/10.1159/000261880>
- Diehl, R., & Kluender, K. (1989). Reply to commentators. *Ecological Psychology*, 1(2), 195–225. https://doi.org/10.1207/s15326969eco0102_6

- Escudero, P., Boersma, P., Rauber, A., & Bion, R. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America*, 126(3), 1379–1393. <https://doi.org/10.1121/1.3180321>
- Esling, J., Moisik, S., Benner, A., & Crevier-Buchman, L. (2019). *Voice quality: The laryngeal articulator model*. Cambridge University Press. <https://doi.org/10.1017/9781108696555>
- Ewan, W., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, 2(4), 327–335. [https://doi.org/10.1016/S0095-4470\(19\)31302-6](https://doi.org/10.1016/S0095-4470(19)31302-6)
- Fahey, R., & Diehl, R. (1996). The missing fundamental in vowel height perception. *Perception & Psychophysics*, 58(5), 725–733. <https://doi.org/10.3758/BF03213105>
- Fahey, R., Diehl, R., & Traunmüller, H. (1996). Perception of back vowels: Effects of varying F1-F0 bark distance. *Journal of the Acoustical Society of America*, 99(4), 2350–2357. <https://doi.org/10.1121/1.415422>
- Fischer-Jørgensen, E. (1990). Intrinsic f0 in tense and lax vowels with special reference to german. *Phonetica*, 47(3–4), 99–140. <https://doi.org/10.1159/000261858>
- Garellek, M., & Esposito, C. (2023). Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association*, 53(1), 213–232. <https://doi.org/10.1017/S0025100321000104>
- Genette, J., Gillis, S., & Verhoeven, J. (2025). Intrinsic vowel fundamental frequency in children with and without hearing impairment. *Folia Phoniatica et Logopaedica*, 77(4), 347–361. <https://doi.org/10.1159/000543426>
- Genette, J., Rivera-Espejo, J., Gillis, S., & Verhoeven, J. (2023). Determining spectral stability in vowels: A comparison and assessment of different metrics. *Speech Communication*, 154, 102984. <https://doi.org/10.1016/j.specom.2023.102984>
- Gregory, A. M. (2013). *Laryngeal aspects of infant language acquisition* [Doctoral dissertation, La Trobe University].
- Gussenhoven, C. (1992). Dutch. *Journal of the International Phonetic Association*, 22(1-2), 45–47. <https://doi.org/10.1017/S002510030000459X>
- Gussenhoven, C. (2007). Wat is de beste transcriptie voor het nederlands? *Nederlandse Taalkunde*, 12, 331–350.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396. <https://doi.org/10.1006/jpho.2000.0121>
- Hombert, J., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language*, 37–58. <https://doi.org/10.2307/412518>
- Honda, K. (1983). Relationship between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research*, SR, 73, 269–282.
- Hoole, P., & Honda, K. (2011, July). Automaticity vs. feature-enhancement in the control of segmental F0. In G. Clements & R. R. (Eds.), *Where do phonological features come from?* (pp. 131–171). John Benjamins Publishing Company. <https://doi.org/10.1075/lfab.6.06hoo>
- Hoole, P., Honda, K., Murano, E., Fuchs, S., & Pape, D. (2004). Cricothyroid activity in consonant voicing and vowel intrinsic pitch. *Proceedings of the Conference on Voice Physiology and Biomechanics*.

- Idemaru, K., & Holt, L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *Journal of the Acoustical Society of America*, 133(6), 4232–4246. <https://doi.org/10.1121/1.4802905>
- Jadoul, Y., Thompson, B., & Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15. <https://doi.org/10.1016/j.wocn.2018.07.001>
- Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT Press.
- Johnson, P., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1, 57–93.
- Kapnoula, E., Winn, M., Kong, E., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, 43(9), 1594. <https://doi.org/10.1037/xhp0000410>
- Keelan, E., Lai, C., & Zechner, K. (2011). The importance of optimal parameter setting for pitch extraction. <https://doi.org/10.1121/1.3609833>
- Kingston, J. (1991). Integrating articulations in the perception of vowel height. *Phonetica*, 48(2–4), 149–179. <https://doi.org/10.1159/000261882>
- Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language*, 70(3), 419–454. <https://doi.org/10.2307/416481>
- Kingston, J., & Diehl, R. (1995). Intermediate properties in the perception of distinctive feature values. *Papers in Laboratory Phonology*, 4, 7–27. <https://doi.org/10.1017/CBO9780511554315.002>
- Kingston, J., Diehl, R., Kirk, C., & Castleman, W. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36(1), 28–54. <https://doi.org/10.1016/j.wocn.2007.02.001>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Wiley.
- Laver, J. (1994). *Principles of phonetics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139166621>
- Lindau, M. (1975). *[Features] for vowels* (UCLA Working Papers in Phonetics, Vol. 30). University of California.
- Lindau, M. (1978). Vowel features. *Language*, 54(3), 541–563. <https://doi.org/10.2307/412786>
- Lisker, L. (1978). In qualified defense of VOT. *Language and Speech*, 21(4), 375–383. <https://doi.org/10.1177/002383097802100413>
- Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1), 3–11. <https://doi.org/10.1177/002383098602900102>
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422. <https://doi.org/10.1080/00437956.1964.11659830>

- Lobanov, B. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49(2B), 606–608. <https://doi.org/10.1121/1.1912396>
- Long, J., & Freese, J. (2001). *Regression models for categorical dependent variables using stata*. Stata Press.
- MacWhinney, B. (2000). The CHILDES project. *Computational Linguistics*, 26(4), 657–657. <https://doi.org/10.1162/coli.2000.26.4.657>
- Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *Journal of the Acoustical Society of America*, 118(3), 1730–1741. <https://doi.org/10.1121/1.1979451>
- McGowan, R., Nittrouer, S., & Manning, C. (2004). Development of [ɪ] in young, Midwestern, American children. *Journal of the Acoustical Society of America*, 115(2), 871–884. <https://doi.org/10.1121/1.1642624>
- Miller, R. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, 25(1), 114–121. <https://doi.org/10.1121/1.1906983>
- Mize, T. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, 6, 81–117. <https://doi.org/10.15195/v6.a4>
- Molemans, I. (2011). *Sounds like babbling. A longitudinal investigation of aspects of the prelexical speech repertoire in young children acquiring Dutch: Normally hearing children and hearing-impaired children with a cochlear implant* [Doctoral dissertation, University of Antwerp].
- Montgomery, D., & Peck, E. (1992). *Introduction to linear regression analysis*. John Wiley & Sons.
- Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20(3), 351–382. [https://doi.org/10.1016/S0095-4470\(19\)30639-4](https://doi.org/10.1016/S0095-4470(19)30639-4)
- Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, 97(1), 520–530. <https://doi.org/10.1121/1.412278>
- Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, 112(2), 711–719. <https://doi.org/10.1121/1.1496082>
- Nittrouer, S., Lowenstein, J., & Tarr, E. (2013). Amplitude rise time does not cue the /ba-/wa/ contrast for adults or children. *Journal of Speech, Language, and Hearing Research*, 56(2), 427–440. [https://doi.org/10.1044/1092-4388\(2012\)12-0075](https://doi.org/10.1044/1092-4388(2012)12-0075)
- Nittrouer, S., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America*, 94, 1865–1865. <https://doi.org/10.1121/1.407649>
- Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Puggaard-Rode, R. (2022). Analyzing time-varying spectral characteristics of speech with function-on-scalar regression. *Journal of Phonetics*, 95, 101191. <https://doi.org/10.1016/j.wocn.2022.101191>

- R Development Core Team. (2022). *R: A Language and Environment for Statistical Computing* [Software manual]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Repp, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81–110. <https://doi.org/10.1037/0033-2909.92.1.81>
- Rietveld, T., & Heuven, V. (2009). *Algemene fonetiek*. Coutinho.
- Rossi, M., & Autesserre, D. (1981). Movements of the hyoid and the larynx and the intrinsic frequency of vowels. *Journal of Phonetics*, 9(2), 233–249. [https://doi.org/10.1016/S0095-4470\(19\)30938-6](https://doi.org/10.1016/S0095-4470(19)30938-6)
- Rowe, M., Raudenbush, S., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, 83(2), 508–525. <https://doi.org/10.1111/j.1467-8624.2011.01710.x>
- Samuel, A., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218. <https://doi.org/10.3758/APP.71.6.1207>
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183–204. <https://doi.org/10.1016/j.wocn.2015.07.003>
- Schertz, J., & Clare, E. (2020). Phonetic cue weighting in perception and production. *Cognitive Science*, 11(2), e1521. <https://doi.org/10.1002/wcs.1521>
- Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology*, 10(1). <https://doi.org/10.5334/labphon.67>
- Shadle, C., Nam, H., & Whalen, D. (2014). Accuracy of six techniques for measuring formants in isolated words. *Journal of the Acoustical Society of America*, 135(4), 2426–2426. <https://doi.org/10.1121/1.4878067>
- Spiller, S., Fitzsimons, G., Lynch, J., & McClelland, G. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, 50(2), 277–288. <https://doi.org/10.1509/jmr.12.0420>
- Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17(1–2), 3–45. [https://doi.org/10.1016/S0095-4470\(19\)31520-7](https://doi.org/10.1016/S0095-4470(19)31520-7)
- Stevens, K., Keyser, S., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426–463). Erlbaum.
- Strange, W., Weber, A., Levy, E., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America*, 122(2), 1111–1129. <https://doi.org/10.1121/1.2749716>
- Sussman, J. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *Journal of the Acoustical Society of America*, 109(3), 1173–1180. <https://doi.org/10.1121/1.1349428>

- Syrdal, A., & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79(4), 1086–1100. <https://doi.org/10.1121/1.393381>
- Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69(5), 1465–1475. <https://doi.org/10.1121/1.385780>
- Van den Berg, R. (2012). *Syllables inside out* [Doctoral dissertation, University of Antwerp].
- Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, 39(2), 168–177. <https://doi.org/10.1016/j.wocn.2011.02.007>
- Van Severen, L. (2011). *A large-scale longitudinal survey of consonant development in toddlers' spontaneous speech* [Doctoral dissertation, University of Antwerp].
- Vanormelingen, L., Maeyer, S., & Gillis, S. (2016). A comparison of maternal and child language in normally-hearing and hearing-impaired children with cochlear implants. *Language, Interaction and Acquisition*, 7(2), 145–179. <https://doi.org/10.1075/lia.7.2.01van>
- Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association*, 35(2), 243–247. <https://doi.org/10.1017/S0025100305002173>
- Verhoeven, J., & Connell, B. (2024). Intrinsic vowel pitch in Hamont Dutch: Evidence for f0 reduction in the lower pitch range. *Journal of the International Phonetic Association*, 54(1), 108–125. <https://doi.org/10.1017/S0025100323000129>
- Vihman, M., & McCune, L. (1994). When is a word a word? *Journal of Child Language*, 21(3), 517–542. <https://doi.org/10.1017/S0305000900009442>
- Walley, A., & Carrell, T. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73(3), 1011–1022. <https://doi.org/10.1121/1.389149>
- Whalen, D., & Levitt, A. (1995). The universality of intrinsic f0 of vowels. *Journal of Phonetics*, 23(3), 349–366. [https://doi.org/10.1016/S0095-4470\(95\)80165-0](https://doi.org/10.1016/S0095-4470(95)80165-0)
- Whalen, D., Levitt, A., Hsiao, P., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97(4), 2533–2539. <https://doi.org/10.1121/1.411973>
- Zink, I., & Lejaegere, M. (2002). *N-CDIs: Lijsten voor communicatieve ontwikkeling [NCDIs: Lists for communicative development]*. Acco.

