University of Amsterdam

Hearing Quantity: Durational and Pitch Cue Use in Estonian Children’s Perception of the Quantity 2 and 3 Contrast

BA Thesis Linguistics

Meeli Lepik, 12121169
6-22-2026

Supervised by Cesko Voeten

Table of Contents

Abstract

Estonian distinguishes three phonemic quantity degrees – short (Q1), long (Q2) and overlong (Q3). While the adult perception of this contrast is well documented, its acquisition in childhood remains poorly understood. This thesis investigates how first-language Estonian speaking children aged 7 to 13 perceive the Q2/Q3 distinction and whether they rely on durational or pitch cues. Using an AXB forced-choice task, 121 children from southwestern Estonia judged stimuli drawn from stressed-vowel duration and pitch continua. Responses were analyzed with mixed-effects logistic regression. Children showed partial, developing sensitivity towards durational cues, though never reaching adult-like categorical sharpness. In contrast, pitch-contour alignment produced no reliable discrimination at any age. Older, formally instructed children showed a modestly steep durational slope, yet, unexpectedly, a stronger Q3 bias in pitch condition that younger children. These findings suggest that durational perception of Estonian quantity consolidates gradually across the school years, whereas the tonal component remains unacquired. The results are consistent with the hypothesis that pitch-based quantity marking is a recent and still-consolidating development in Estonian prosody.

1. Introduction

Estonian, a Finno-Ugric language spoken by approximately 1.1 million people (Winter 2022), is distinctive for its ternary system of phonemic quantities for both vowels and consonants. These quantities are referred to as short (Q1), long (Q2), and overlong (Q3) and play a pivotal role in distinguishing meaning both lexically and grammatically (see examples 1 & 2).

<u>Example 1. Estonian minimal triplet: lexical distinction.</u>

a) vere /ˈvere/ = blood.GEN.SG

b) veere /ˈveːre/ = edge.GEN.SG

c) veere /ˈveːːre/ = roll.2SG.IMP

<u>Example 2. Estonian minimal triplet: grammatical distinction.</u>

a) (sada /ˈsɑdɑ/ = hundred.NOM)

b) saada /ˈsɑːdɑ/ = to get/send.INF

c) saada /ˈsɑːːdɑ/ = to get/send.2SG.IMP

Vowel quantity in Estonian refers to the segmental length distinction, whose phonetic realisation is conditioned by the structure of the foot. While most languages of the world operate within a binary system of short versus long vowels, Estonian's ternary system is rather rare, as it introduces a third category of overlong vowels. These types of vowels are orthographically indistinguishable from their long counterparts but with distinctive differences in duration and pitch (Q2 and Q3 are distinguished not only by vowel duration but also by the alignment of the F0 contour within the stressed syllables – see section 2.3). Among the three, Quantity 2 and Quantity 3 stand out as the more abstract and less universally understood components of the system, due to the close interplay of duration, stress and pitch.

1.1 Background of Vowel Quantities in Estonian

The Estonian language's vowel quantity system is rooted in its Southern Finnic heritage. Earlier studies have recognized the contrast between long and overlong vowels in Estonian, features that diverge from closely related languages such as Finnish and Livonian but do not exclude language contact with the geographically close Baltic languages (Pajusalu 2009: 109-121). The three-way quantity distinction is most prominent in stressed syllables, where prosodic differences are realized as variations in segmental duration as well as distinctive pitch contour. While Q2 and Q3 vowels are phonologically long, their durational difference in natural speech is speaker-variable rather than categorically distinct, which is why the pitch cue has become important for disambiguation.

According to Pajusalu (2012: 206), the third, overlong quantity displays a distinctively high beginning pitch contour with a fall and a low ending, while its long (Q2) counterpart remains relatively even in pitch throughout the duration of the vowel. The quantity contrast of vowels only occurs in the primary stressed syllable (in the case of Estonian, the very first syllable of the word), in unstressed syllables the contrast is neutralized - the durations are inversed (Q1 in primary syllable leading to a longer vowel duration in the second syllable and Q2 or Q3 in primary syllable leading to a shorter vowel duration in the secondary syllable) (Asu & Teras 2009: 370).

The historical development of Estonian's quantity system is believed to result from diachronic changes in syllable structure and stress patterns. Early studies, such as those by Johannes Aavik and Jakob Hurt, both renowned 19th-century Estonian literati, illustrated the functional importance of these distinctions in spoken and written Estonian through translation, poetry, and creative wordplay. More recent research by Lippus (2011), incorporating acoustic phonetics and computational modelling, suggests that Estonian is in the process of typological change from a quantity language towards a pitch accent language, where the majority of speakers prefer pitch as the decisive factor when distinguishing between Q2 and Q3 sounds. Pajusalu (2009) further suggests that speakers from the southeast of Estonia can distinguish between the two longer lengths just based on the duration of the elements, whereas speakers from western and central Estonia prefer the fundamental frequency contour as their distinguishable factor, possibly due to historic language contact (Swedish contact in western Estonia and Russian in the east).

1.2 Research Objectives and Questions

This thesis aims to explore the perceptual strategies of native Estonian-speaking children aged 7-13 (a group whose prosodic and phonological and orthographic awareness is still developing) in relation to Quantity 2 and Quantity 3 vowels. This age range was selected as the formal education on the quantity system commences at around age 9 (Riigi Teataja, 2002), meaning children aged 7-9 in this sample have not yet received explicit teaching about the distinction, while children older than age 10 have. Whether instruction improves the perceptual representation of these categories, and whether duration or pitch is the more accessible cue at each age group, are the core questions of this study.

The research questions guiding this study include:

1) How well do Estonian L1-speaking children aged 7-13 perceive differences between long and overlong vowel quantities?
2) How do age and formal education affect their perception?
3) To what degree does pitch contour seem necessary for the participants to distinguish between the quantities?

The investigation of Q2 and Q3 vowels reveals how listeners perceive and internalize subtle prosodic cues, leading possibly to a better understanding of synchronic and diachronic language change. Measuring the effect of this process in children, who are at various stages of quantity acquisition, is an important indicator of the current ongoings within the language typology change.

Furthermore, this research has practical implications for language educators. A perceptually informed view of vowel quantity acquisition can lead to more effective and efficient teaching methods, helping not only non-native learners but also children and adults with language-related disabilities towards fluency. A better understanding of pitch contour and durational effects in relation to perception can also benefit speech recognition and language processing applications.

1.3 Structure of the Thesis

This thesis is organized into the following chapters: Chapter 2 reviews the relevant previous literature on Estonian vowel quantities, focusing on the phonetic and phonological properties of Q2 and Q3. Chapter 3 outlines the methodology, stimuli, and the participants, including data collection and statistical techniques. Chapter 4 presents the findings of the study, highlighting the key differences between Q2 and Q3 in perception and relevant variables. Finally, Chapter 5 discusses the limitations, the implications of these findings for theoretical linguistics and practical applications, concluding with suggestions for future research.

2. Literature review

2.1 The Estonian phonological system

Estonian has a rich sound system that forms the backdrop against which the quantity system operates. The vowel inventory consists of nine phonemes: /i, y, e, ø, æ, ɑ, o, ɤ, u/, distributed across three heights (high: /i, y, u/; mid: /e, ø, ɤ, o/; low: /æ, ɑ/) and opposing front versus back and rounded versus unrounded articulations (Asu & Teras 2009: 367–369). Of the nine vowels, /y, ø, ɤ, æ/ are largely restricted to word-initial (primary stressed) syllables, while /i, e, ɑ, o, u/ may also occur in non-initial positions (Asu & Teras 2009: 369). Estonian, furthermore, has a diphthong inventory of 36 combinations, 28 of which are considered native to the language, with all nine vowels capable of appearing as the first element but only /ɑ, e, i, o, u/ as the second (Asu & Teras 2009: 369).

The consonant inventory is comparatively lean, with the fricatives /f/ and /ʃ/ occurring only in loanwords, and a notable feature of non-phonemic palatalization occurring before front vowels for all word-internal and word-final alveolar consonants except /r/ (Asu & Teras 2009: 370). As for the quantity system, most consonants, like the vowels, participate in the three-way length distinction (Asu & Teras 2009: 370).

Stress in Estonian is fixed to the first syllable of a word, which is also the location of the quantity contrast. The phonological unit within which quantities are defined is the disyllabic foot: quantity degrees are properties of the foot as a whole, not simply of individual segments, and are realized through the relationship between the primary stressed syllable and the following unstressed syllable (Lippus et al. 2013: 17; Lehiste 1960). This foot-level view is now broadly accepted in the literature, superseding earlier analyses that treated quantity as a property of individual segments alone (Prillop 2013).

2.2 The quantity system: acoustic and phonological properties

The three-way quantity contrast operates primarily on the primary stressed (first) syllable, where vowels and consonants alike may be short (Q1), long (Q2), or overlong (Q3). The opposition is realized through variation in segmental duration, pitch contour, and vowel quality simultaneously (Lippus et al. 2013). The Q1/Q2 distinction is primarily durational: the stressed-syllable rhyme is longer in Q2 than in Q1, and the duration ratio of stressed-to-unstressed syllables rises accordingly (Lehiste 1960; Eek 1974). For the Q2/Q3 distinction, however, duration alone is insufficient: while Q3 feet do show a longer stressed syllable rhyme and a higher S1/S2 duration ratio than Q2, the difference is more variable in natural speech, and pitch becomes an essential additional cue (Lippus et al. 2013; Lippus, Pajusalu & Allik 2009).

Vowel quality also relates to the quantity in a systematic way. In spontaneous speech, the vowel space of the stressed syllable contracts slightly under Q1 conditions, while Q3 vowels tend to show more extreme formant values (Lippus et al. 2013). Furthermore, the quality of the unstressed syllable vowel varies with the quantity of the foot: in Q2, the unstressed vowel is typically longer and may be closer to a full vowel, whereas in Q3, it is shorter relative to the stressed syllable (Eek & Meister 1998; Lippus et al. 2013). This multi-dimensional nature of the contrast - encompassing duration, pitch, and quality simultaneously - is what makes it typologically unusual and perceptually complex.

2.3 Pitch contours and their role in quantity perception

The tonal component of the Estonian quantity system has attracted continuous research attention, as it is the primary acoustic feature differentiating Q2 from Q3 for most native speakers. In Q1 and Q2, the fundamental frequency (F0) contour is characterized by a step-down between the end of the first stressed syllable nucleus and the onset of the second unstressed syllable: the pitch is relatively flat or gently declining during the stressed vowel, then falls across the syllable boundary (Lippus et al. 2013: 18; Asu & Teras 2009). In Q3, by contrast, the F0 peaks earlier (within the first half of the stressed vowel) so that a clear pitch fall takes place during the stressed syllable itself, leaving both a high and a low plateau within V1 (Lippus, Pajusalu & Allik 2011: 217). The stressed syllable of a Q3 foot is therefore associated with an H*+L pitch accent, while Q1 and Q2 have a later or boundary-aligned F0 peak.

The perceptual weight of this tonal cue has been demonstrated across multiple experimental studies. Eek (1980) showed that perception tests using stimuli with manipulated segment durations revealed that Q3 could only be perceived reliably when the F0 fall was located within the first syllable, rather than at its boundary. Lippus, Pajusalu & Allik (2009) confirmed this finding with a larger stimulus set, reporting that native Estonian L1 listeners failed to perceive Q3 when pitch cues were consistent with

Q1 or Q2 - even when the segmental duration ratios were typical of Q3. Conversely, it was later shown that altering only the pitch contour of a Q2 word (without any change to segment durations) was sufficient to induce Q3 perception in listeners (Lippus, Pajusalu & Allik 2011: 218). This demonstrates that pitch is not merely a secondary cue but can under certain conditions override durational evidence in determining perceived quantity.

The precise location of the F0 turning point (TP) within the stressed vowel is the critical variable. Lippus, Pajusalu & Allik (2011: 217) found that the pitch contour triggering Q3 perception specifically requires the fall to occur in the middle of V1, producing both a high and a low plateau within the stressed syllable. A steadily falling F0 throughout V1, or a fall located early or late in V1, favours Q2 rather than Q3 perception. This narrow perceptual boundary between Q2- and Q3-consistent pitch contours makes the cue cognitively demanding for listeners who have not fully internalized it. Dialect variation adds a further layer of complexity as listeners from western and central Estonia rely more heavily on pitch for Q2 and Q3 discrimination, while speakers from southeastern Estonia remain more sensitive to duration alone (Lippus & Pajusalu 2009; Pajusalu 2009). As noted by Lippus (2011), this regional difference is consistent with the hypothesis that Estonian is undergoing a typological shift from a quantity language toward a pitch-accent language, with the change proceeding at different rates across dialect areas.

An important orthographic consequence of the tonal distinction is that Q2 and Q3 are not differentiated in Estonian writing - both are represented by double vowel or double consonant letters (e.g., the word pair *saada* ‘to get/send.INF’ [Q2] and *saada* ‘to get/send.IMP’ [Q3] are written identically) (Meister & Meister 2013; Prillop 2013). This orthographic similarity means that the tonal and durational cues cannot be recovered from the written form of the word.

The general relationship between pitch and duration in quantity is best characterized as an enhancement relationship, rather than a trading one. Both cues are correlated in natural speech (greater duration and earlier F0 fall both signal Q3), but their relative perceptual weighting varies across dialect regions.

2.4 Acquisition of vowel quantities in L1 Estonian children

The acquisition of quantity contrasts by children learning Estonian as their first language is embedded within a broader trajectory of phonological development. By the age of around three years, Estonian-acquiring children typically produce the syllable structures and quantity of shorter words correctly, though longer polysyllabic words may be simplified (Lahtein, Mailend & Padrik 2021). The quantity system is among the earliest prosodic features to be acquired, likely because foot-level prosodic templates are highly salient in ambient input from the earliest stages of language development (Vihman, Nakai & DePaolis 2006).

Longitudinal research on early phonological development in related Finnic languages, particularly Finnish, provides a useful comparative framework. Kunnari, Nakai & Vihman (2001) showed that Finnish-acquiring children initially overuse long consonants in their early productions, but that this overuse is gradually corrected as ambient language-specific patterns are internalized. Given that Estonian’s quantity system is more complex than Finnish's (three degrees versus two, with a tonal component), it is reasonable to expect that full adult-like mastery of the Q2/Q3 distinction takes longer to emerge. Cross-linguistic work by Vihman & Kunnari (2011) suggests that phonological knowledge is built step-by-step from exposure to ambient language, which for Estonian children would mean that the characteristic pitch contour of Q3 must be abstracted from the input over the course of extended phonological development.

Neurophysiological evidence for the developmental trajectory of quantity perception in Estonian children comes from EEG studies using the Mismatch Response (MMR). Lippus and colleagues' longitudinal work with preschool-aged children (mean age approximately 5–6 years) demonstrated a notable developmental shift: in the duration-change condition, the MMR transitions from a positive polarity at around age 5;8 to a negative polarity at around age 6;8, with this transition being less pronounced in children with Developmental Language Disorder (DLD) compared to typically developing peers (Themas et al., 2025).

This shift in MMR polarity is interpreted as reflecting a maturational change in auditory cortical processing. The earlier positive MMR is thought to reflect more immature neural encoding, while the negative MMR seen in older children and adults reflects a more adult-like detection of phonological deviance. The fact that this shift is delayed in DLD children underscores that quantity perception is not simply a perceptual given but a developmentally acquired sensitivity that can be disrupted by language processing difficulties (Lippus et al. 2019).

Lahtein, Mailend & Padrik (2021) argue from a psycholinguistic perspective that quantity is a lexically stored property of word forms in Estonian - its planning originates at the phonological level of speech production, and impairments at that level manifest in children's speech as errors in the carrier segment of quantity (i.e., the vowel or consonant through which the quantity is expressed) or in the durational proportions characteristic of each degree. Since stress is regular in Estonian (always word-initial), quantity errors are potentially diagnostically informative in clinical assessment of Estonian-speaking children, as they would reflect specifically phonological-level deficits rather than motor execution failures.

2.5 Quantity acquisition and the broader phonological system

The acquisition of the quantity system does not happen in isolation from the rest of the phonological system. Since quantity is a property of the entire disyllabic foot, its mastery is intertwined with the acquisition of syllable structure, stress assignment and the segmental inventory. For Estonian children, the large vowel inventory (nine phonemes) and the restricted distribution of certain vowels (/y, ø, ɤ, æ/) mean that the segmental and prosodic levels of the phonological system are closely linked. Acquiring the knowledge of where vowels occur also entails acquiring the foot structure within which quantity is relevant.

The interaction of quantity with morphology is particularly significant for acquisition. Estonian employs the quantity contrast extensively in its morphological system: the same segmental string can represent different grammatical forms depending on the quantity degree of the foot (see Examples 1 and 2 in Chapter 1). This means that children acquiring Estonian must learn to use prosodic features for grammatical purposes from an early stage, making the process different from languages where morphological distinctions are signalled by segmental suffixes alone.

Argus (2008) documents that Estonian children begin acquiring the morphological case system around the age of two and that the interaction of quantity with case forms presents challenges throughout the preschool years. The acquisition of morphologically conditioned quantity alternations (where the quantity of a stem changes depending on the case or tense form required) requires the child to simultaneously master the prosodic form and its grammatical function, adding to the general cognitive load of quantity acquisition.

In the context of general prosodic development, Estonian children's quantity acquisition can be situated within what Vihman, Nakai & DePaolis (2006) call the process of "getting the rhythm right": the progressive alignment of early productions with the ambient language's specific prosodic templates. The foot structure of Estonian - with its fixed initial stress and the quantity contrast realized within the disyllabic foot - provides a highly salient template for early word learning. Cross-linguistically, children acquiring quantity languages tend to show early sensitivity to durational distinctions in their ambient input (Kunnari, Nakai & Vihman 2001), but the specifically tonal component of the Estonian Q2/Q3 distinction adds a layer of complexity not found in Finnish or other closely related quantity languages.

2.6 The Role of Formal Education in Quantity Acquisition

Formal schooling plays a distinctive role in Estonian children's awareness of the quantity system. Unlike most phonological contrasts, which are fully mastered before school age and need not be taught explicitly, the Q2/Q3 distinction presents a challenge that formal instruction directly addresses. The Estonian school curriculum introduces quantity as a named linguistic concept at the begin of the third grade, approximately at age nine (Riigi Teataja 2002). The reasoning for this timing is partly orthographic: while Q1 and Q2 map onto single versus double spelling of vowels and most consonants in a fairly transparent way, the Q2/Q3 distinction is not marked in writing at all, requiring explicit metalinguistic awareness for spelling and reading comprehension of grammatically ambiguous forms.

The Estonian orthographic system is generally regarded as transparent (a high degree of grapheme-to-phoneme correspondence) but the Q2/Q3 neutralization in spelling creates a specific opacity precisely at the point where the prosodic system is most complex.

Soodla et al. (2015) found that Estonian and Finnish children, who both learn to read in transparent orthographies, develop decoding skills relatively rapidly (consistent with the cross-linguistic finding that orthographic transparency supports early reading acquisition). However, the quantity-orthography mapping in Estonian is not fully transparent at the Q2/Q3 level, and this residual opacity may contribute to the observed continuation of instructional attention throughout basic education.

The metalinguistic dimension of quantity instruction is significant. Before formal schooling, Estonian children use quantity correctly in productive speech and perceive it in input without necessarily being able to articulate the contrast as a linguistic object. The introduction of the quantities in the classroom asks children to shift from implicit phonological knowledge to explicit metalinguistic knowledge - a transition that has been shown more broadly to be cognitively demanding and age-sensitive (Bialystok 2001). For children aged 7–9 who may not yet have fully consolidated the perceptual representation of the Q2/Q3 contrast, this metalinguistic demand may precede the consolidation of the underlying perceptual category, which, in turn, could explain variability in performance on Q2/Q3 perception tasks in the age range examined in this study.

Research on related processes in Finnish shows that children's phonological awareness of quantity-related durational contrasts develops across the early school years and is intimately linked with literacy acquisition (Soodla et al. 2015). In Estonian, where the system is additionally modulated by pitch, one would expect this development to be lengthy and to show more variance across children at the ages most relevant to this study (7–13 years). The educational intervention that begins at grade 3 thus falls at a point where some children may still be in the process of perceptual consolidation of the Q2/Q3 contrast, making the school years a critical window for studying the interaction between implicit phonological knowledge and explicitly taught metalinguistic categories.

It should also be noted that explicit instruction on quantity in Estonian schools is not simply perceptual training, but rather embedded within literacy instruction, where children are taught to identify the correct quantity degree of written words as a means of correct pronunciation and grammar. This situates quantity awareness within a reading-and-writing framework rather than a purely phonological one.

Orthographic depth research (Viise, Richards & Pandis 2011, cited in Soodla et al. 2015) suggests that Estonian's generally shallow orthography supports efficient literacy acquisition but that the non-marking of the Q2/Q3 distinction in writing creates a localized depth that is specifically addressed by the school curriculum. Children who receive this instruction may develop sharper metalinguistic representations of the contrast than pre-school children, even if their underlying perceptual sensitivity was already in place - a dissociation that would manifest as improved consistency and confidence in forced-choice quantity discrimination tasks following the onset of formal instruction. In total, four possibilities can be distinguished:

1) The child has not acquired the system yet (no perceptual representation of Q2/Q3), resulting in at chance performance across the continua.
2) The child has implicit perceptual sensitivity but no metalinguistic awareness, resulting in above chance discrimination.
3) The child has both implicit perception and metalinguistic awareness, resulting in sharper category boundaries (scoring more correctly in a discrimination task).
4) The child has metalinguistic awareness without a consolidated perceptual representation (possible for children, who have received formal instruction before their phonological system is ready). Resulgint in a Q3-biased response at all continuum steps.

The remainder of this thesis discusses an AXB task to discriminate between these possibilities.

3. Methodology

3.1 Participants

This study recruited 128 L1 Estonian-speaking children aged 7 to 13 years old (see age distribution of the participants in Table 1). The participants' parents or legal guardians were asked to provide consent and confirm no existence of hearing, visual, or other language-related disabilities, as well as Estonian being used as a home language.

Table 1. *Age and years of quantity-related education distribution of the participants*

Age	Number of participants	Years of quantity-related education
7	8	0
8	19	0
9	13	0
10	30	1
11	34	2
12	15	3
13	9	4

All participants were enrolled in formal education and were recruited within the city of Pärnu, as well as a small village Häädemeeste. These two locations were selected primarily for practical reasons. Both locations were familiar to the researcher, which made it feasible to recruit a sufficient number of child participants within the available time. They are both located in southwestern Estonia, in the general dialect region where adult speakers show the most robust pitch-based discrimination (Lippus & Pajusalu 2009; Pajusalu 2009). Recruiting from this area means participants grow up with ambient input from adults who rely predominantly on pitch as the decisive cue.

3.2 Stimuli

The stimuli were recorded by an adult female native Estonian speaker and modified in Praat. The target stimuli (see Table 2) were chosen for the combinatory reason of forming a minimal pair with contrastive meaning, as well as for the presence of vowel /ɑ/, which has a consistently high occurrence rate in most Estonian dialects (Lindström & Pajusalu, 2003). The stimuli were edited in Praat (Boersma et al. 2025) and organized into two separate continua (pitch vs. duration).

Table 2. *Chosen stimuli minimal pair 'kaalu'.*

Quantity	Orthography	IPA	Gloss
Q2	Kaalu	ˈkɑːlu	scale.GEN.SG
Q3	Kaalu	ˈkɑːːlu	scale.PART.SG/scale.ILL.SG

The durational stimuli were modified by measuring the length of the vowel /ɑ/ within the Q2 (see figure 1) and Q3 (see figure 2) recordings and calculating their durational difference (around 0.15s). The difference was then divided by 10, resulting in a step length of 15ms. Q3 stimuli were then shortened on a continuum of 10 steps to provide an equal number of targets.

Quantity 3 stimulus was chosen for the modifications for being the more extreme form of the two, as well as for reliability issues: lengthening by durational manipulation risks introducing pitch period interpolation artifacts that could distort the F0 contour. Furthermore, lengthening instead of shortening might risk overshooting into the exaggerated, perceptually unnatural durations. Both Q2 and Q3 stimuli were flattened in pitch to measure purely durational effects.

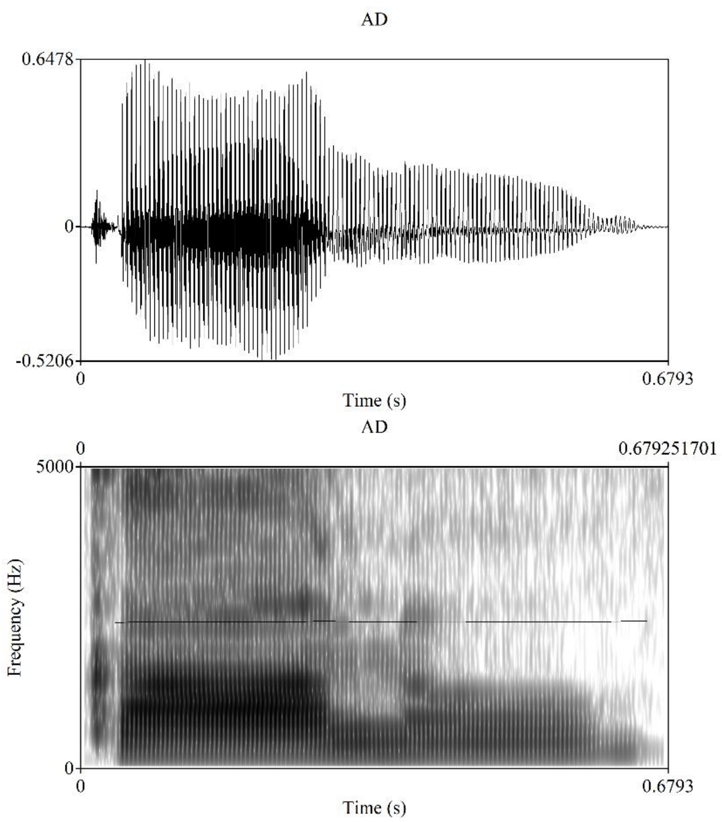

Figure 1. *Reference stimulus A (Quantity 2) flattened in pitch for the duration continuum.*

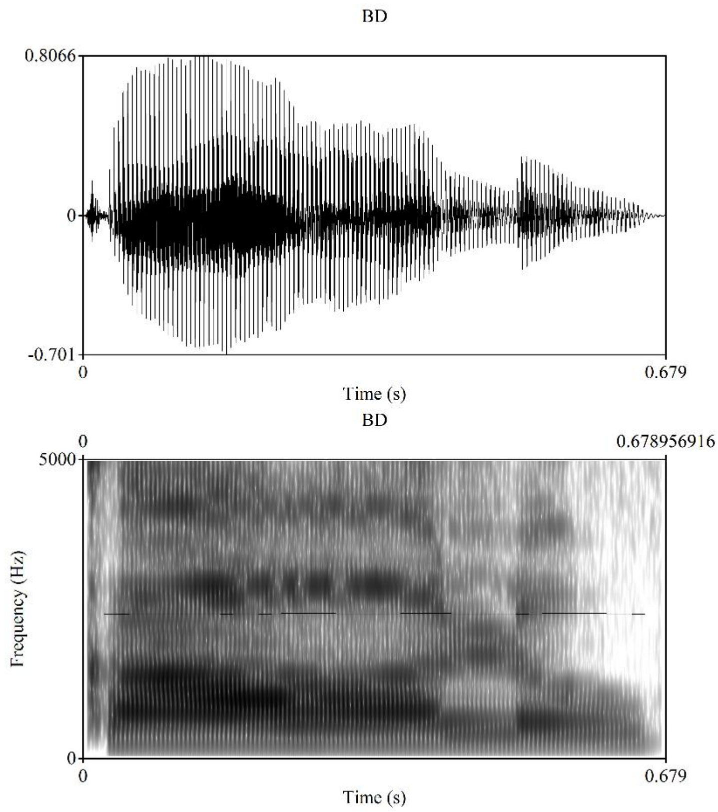

Figure 2. *Reference stimulus B (Quantity 3) flattened in pitch for the duration continuum.*

As for the pitch contour stimuli, the duration of the vowel was kept as-is. As the Q3 pitch contour (see Figure 4) has been observed to begin dropping early in the production of the vowel, and the Q2 pitch contour (see Figure 3) drops just before the beginning of the second syllable, Q3 stimuli were chosen to be modified. The pitch contour was modified manually to mimic most naturalistic-sounding speech with equal intervals, where pitch points were raised to delay the eventual drop (see Figure 5), resulting in a continuum of 10 pitch stimuli.

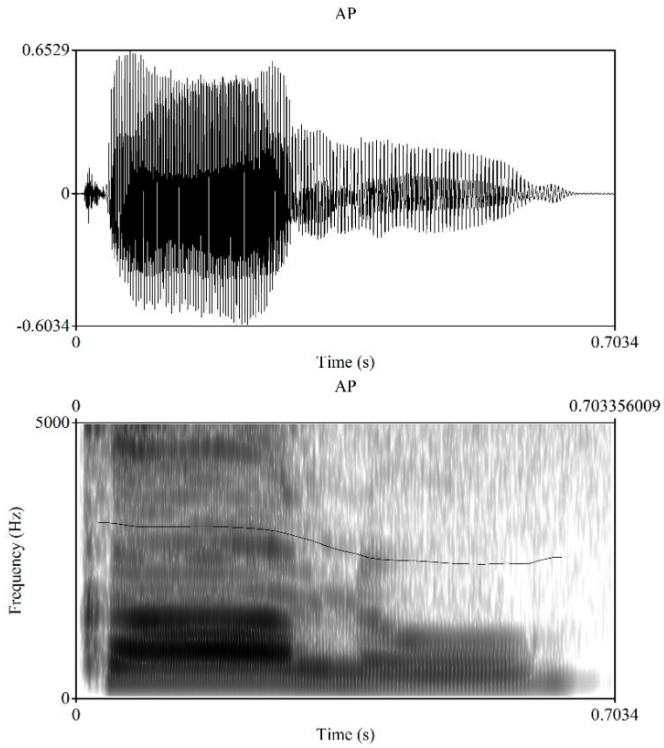

Figure 3. *Reference stimulus A (Quantity 2) in pitch continuum.*

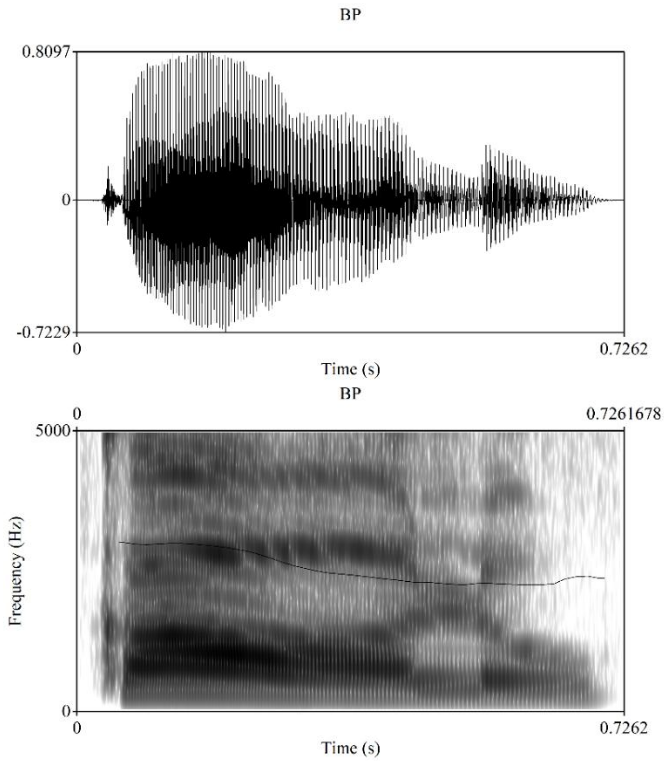

Figure 4. *Reference stimulus B (Quantity 3) in pitch continuum.*

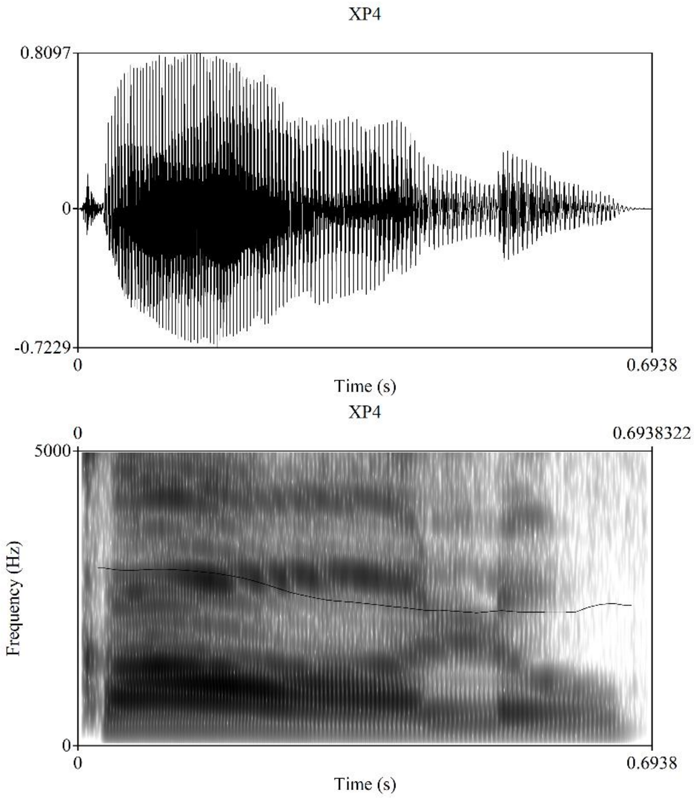

Figure 5. *Modified target stimulus X_4 in pitch continuum.*

Stimuli were arranged in two separate continua, according to their cue type and presented in a randomized order per participant with both continua mixed together.

3.3 Procedure

The experiment consisted of an AXB decision task, where the participants would determine whether the vowel in the word (stimuli X) they hear is more similar to the long (Q2, stimuli A) or the overlong (Q3, stimuli B) vowel. The test was conducted in an online environment, coded by Dirk Vet from University of Amsterdam Speech Labs. The participants were provided with headphones and a computer from their school's computer class, on which they heard the stimuli in either AXB or BXA order and were instructed to press either left or right arrows on a keyboard, matching the order of presentation. Left arrow indicated 'more similar to sound A', right arrow indicated 'more similar to sound B', in BXA order the arrows were reversed (left indicating 'more similar to sound B' and right indicating 'more similar to sound A').

Participants were instructed to answer based on their first instincts, yet not to hurry, as the experiment did not reward those finished first. The experiment began with a short warm-up sequence, where participants could familiarize themselves with the testing environment and control their understanding of the task at hand.

The practice round was then followed by first presenting stimuli in AXB ordering, where the participants heard the target stimulus A (Q2), indicated by a blue circle, then either modified pitch or duration stimuli (indicated by a white circle), and finally the target stimulus B (Q3) indicated by yellow circle. After hearing all three stimuli, both blue and yellow circle appeared simultaneously, during which participants would be able to press either left or right arrow (see Figure 6). There were in total 20 stimuli, combined from both continuums in a randomized order.

The same procedure was immediately followed by BXA ordering, where Q2 and Q3 presenting positions of the reference stimuli were switched, resulting in a total of 40 observations per participant. Presenting both orderings to every child meant any endpoint-priming bias (hearing the Q2 stimuli first in AXB) was captured within each participant rather than confounded between groups and was absorbed by the by-participant random intercept in the mixed-effects model.

Figure 6. *Decision making screen*.

The experiment was conducted at the presence of their class teacher, the length of the experiment was around 7 minutes, with an additional 20-30 minutes per class for instructions, technical setup and troubleshooting. All participants were gifted a freely chosen sweet as a thank you for participation.

3.4 Statistical Analysis

Data were analyzed using mixed-effects logistic regression in R. The outcome variable was the binary forced-choice response (1 = chose Q2, 0 = chose Q3). The fixed effects included: Step (the continuum position, centred at zero), Trial Type (duration vs. pitch), and Age Group (younger vs. older), along with all two-way and three-way interactions.

Both categorical predictors were treatment-coded: duration and younger (aged 7 to 9) served as reference levels for Trial Type and Age Group respectively. The model intercept therefore represents the estimated log-odds of a Q2 response at the continuum midpoint for a younger participant in the duration condition. A by-participant random intercept and random slope for step were included to account for individual differences in baseline response probability and perceptual sensitivity.

Random slopes for Trial Type and Step/Trial Type interaction were not included. The maximal by-participant structure produced a singular fit, indicating that the data cannot support these additional variance components. Item random effects were not included, as the steps formed a fixed set of stimuli presented to every participant rather than a random sample of possible ones.

Model comparisons were conducted using likelihood ratio tests (LRT), in order to evaluate whether the three-way interaction between Step, Trial Type and Age Group significantly improved model fit over a two-way interaction model, as the three-way interaction involves a relatively small effect that would otherwise be difficult to justify retaining.

The full fixed-effects regression coefficient table (see Table 3) is reported with each effect as a coefficient (β), its odds ratio (OR), 95% confidence interval (CI) and z- and p-values. Furthermore, estimated marginal means are reported for interpretability only, to express selected effects as model-predicted probabilities of a Q2 response.

Table 3. *Fixed-effects coefficients from the mixed-effects logistic regression.*

Predictor	**β (SE)**	**z**	**OR**	**95% CI (OR)**	***p***
Intercept	−0.229 (0.127)	−1.80	1 : 1.26	[1 : 1.61, 1.02 : 1]	.071
Step (Duration slope)	0.123 (0.033)	3.71	1.13 : 1	[1.06 : 1, 1.21 : 1]	< .001*
Trial Type: Pitch	−0.878 (0.114)	−7.74	1 : 2.41	[1 : 3.00, 1 : 1.93]	< .001*
Age Group: Older (10–13)	−0.100 (0.156)	−0.64	1 : 1.10	[1 : 1.50, 1.23 : 1]	.522
Step × Pitch	−0.105 (0.040)	−2.63	1 : 1.11	[1 : 1.20, 1 : 1.03]	.008*
Step × Older	0.083 (0.041)	2.04	1.09 : 1	[1.00 : 1, 1.18 : 1]	.041*
Pitch × Older	−0.395 (0.142)	−2.78	1 : 1.48	[1 : 1.96, 1 : 1.12]	.005*
Step × Pitch × Older	−0.092 (0.050)	−1.84	1 : 1.10	[1 : 1.21, 1.01 : 1]	.066

4. Results

4.1 Participants and data preparation.

The dataset originally consisted of 128 individual participants. During testing, four participants discontinued the task and were excluded from the analysis. Discontinuation was attributed to task fatigue or unwillingness to continue, as reported by their accompanying teacher and observed during testing. During data analysis, one participant sheet was excluded in full because no age value could be recovered. In addition, two sheets produced no valid trials after response cleaning, yielding a final sample of N = 121 participants.

At the trial level, in total 120 observations were excluded from 4959 trials due to recording artefacts: pipe-separated double key presses (e.g. {RIGHT}|{LEFT}, first press retained) and a single alphabetic mis-key. After exclusions, 4839 trials were kept: 2420 for the duration and 2419 for pitch. Each participant contributed a maximum of 40 trials (20 per cue type across AXB and BXA trial types).

Continuum steps were centred prior to analysis. The ten-step continuum ranged from 1 (most Q3-like, X stimulus most similar to unmodified B) to 10 (most Q2-like, stimuli X most similar to unmodified A). This continuum was chosen to create a midpoint, so that the model intercept represents the estimated response probability at the average stimulus value, and all other coefficients represent effects at that point. Responses were recoded such that 1 consistently indicated a preference for the Q2 stimulus (A) and 0 indicated a preference for the Q3 stimulus (B) across both presentation orderings, including accounting for the reversal of key assignments between the AXB and BXA conditions.

4.2 Preliminary check: AXB vs BXA consistency

Figure 7 shows a sanity check comparing AXB and BXA response curves after response harmonization. In the pitch condition the two blocks overlap closely, confirming consistent coding. In the Duration condition, AXB responses are systematically higher than BXA at each step, suggesting a presentation-order effect: hearing A (Q2) first in AXB may prime listeners toward Q2 responses, whereas hearing B (Q3) first in BXA may prime toward Q3.

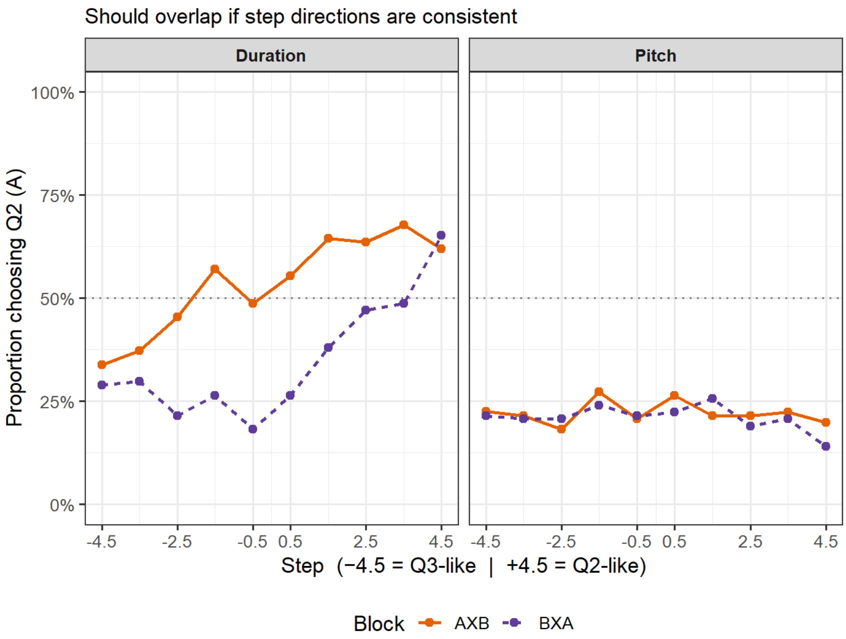

Figure 7. *Sanity check.*

4.3 Overall Q2/Q3 discrimination

Figure 8 presents the percentage of Q2 (A) responses across the ten continuum steps for all 121 participants, separately for Duration and Pitch trials.

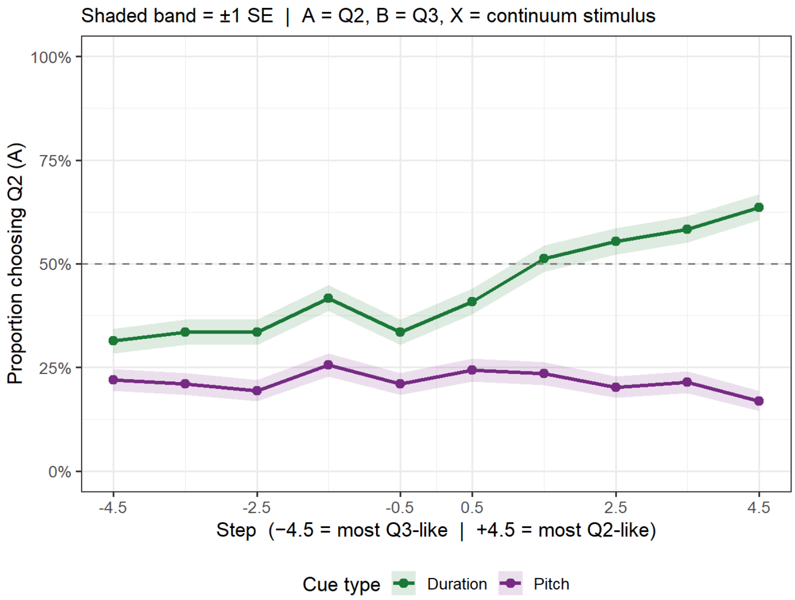

Figure 8. *Percentage of Q2 (A) responses by cue types.*

In the Duration condition, the identification curve rises from approximately 31% at step -4.5 to approximately 64% at step +4.5, crossing the 50% chance line between steps +0.5 and +1.5. This is the expected direction for a sensitive listener, as X becomes more similar to A (Q2), the probability of a Q2 response increases. The curve shows a non-monotonicity near the midpoint (rise at -1.5 followed by a dip at step -0.5), suggesting perceptual sensitivity that is real but not yet adult-like in sharpness.

In the Pitch condition, Q2 response rates are uniformly low (approximately 15-25%) with no distinguishable trend across steps. Children responded Q3 on approximately 75-85% of Pitch trials regardless of the pitch-alignment step. The progressive delay of the F0 turning point from a Q3-like early-falling to a Q2-like late-falling contour does not produce a reliable change in responses.

The estimated marginal mean P(Q2) at the continuum midpoint (Step = 0) was 0.431 (95% CI: [0.394, 0.469]) for duration and 0.205 (95% CI: [0.180, 0.233]) for pitch. The cue type contrast was highly significant (β = -1.08, OR = 2.93, 95% CI [1:3.82; 1:2.27], $z = 15.06$, $p < .0001$). The duration slope (β = 0.165, OR = 1.18, 95% CI: [1.13:1; 1.23:1], $z = 8.14$) was significantly positive ($p < .001$). The Pitch slope (β = 0.014, OR = 1.01:1, 95% CI: [1:1.03; 1.06: 1], $z = 0.62$, $p = .534$) did not differ significantly from zero. The slope contrast was highly significant (Duration minus Pitch = 0.151, $z = 6.00$, $p < .0001$).

4.4 Effects of age and formal education

Participants were divided into a younger group (7-9 years old, n = 40, pre-instruction) and an older group (10-13 years old, n = 81, post-instruction). Figures 9 and 10 show the psychometric curves by age group and overall accuracy by single age year, respectively.

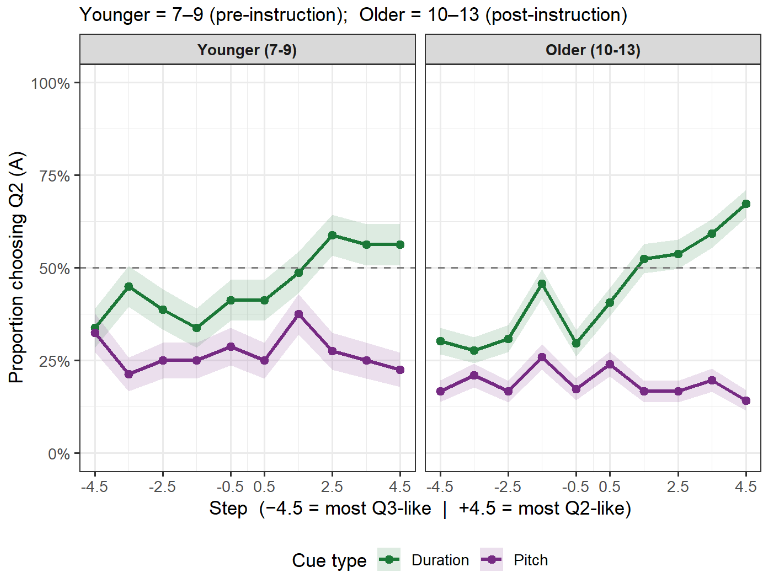

Figure 9. *Psychometric curves by age group and cue type.*

For duration trials, both groups show a positive slope. The older group's curve is steeper (approximately 30% to 68%) versus the younger group (approximately 34% to 59%). The Step x Age Group interaction was significant (OR = 1.09, z = 2.04, p = .041), confirming that the duration slope is reliably steeper in older children, consistent with increased durational sensitivity over the school years.

For pitch trials, both groups show the same flat Q3-biased pattern but the older group (15-20% Q2) is lower than the younger group (20-35% Q2). The Trial Type x Age Group interaction was significant (OR = 0.674, z = -2.78, p = .005). Emmeans contrasts confirmed no midpoint difference for duration (Younger = 0.443, Older = 0.419, p = .522), but a significant difference for pitch (Younger = 0.248, Older = 0.168, OR = 1.64, p = .003). Older children, despite formal instruction, show a stronger and more uniform

Q3 bias in the Pitch condition, opposite to what instruction-driven improvement would predict.

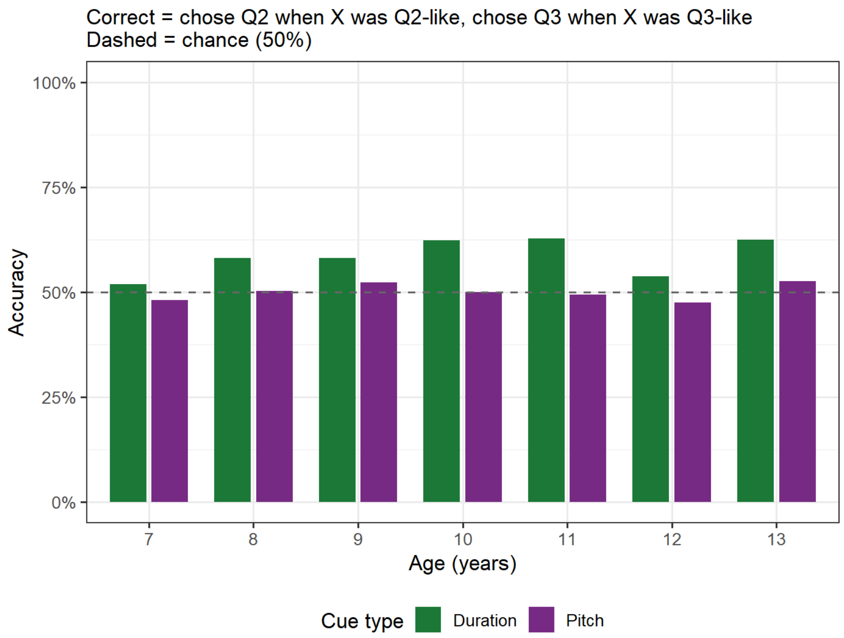

Figure 10. *Accuracy by age year and cue type.*

Through all ages, distinguishing between Q2 and Q3 by pitch remains relatively uniform, around 48-52%. Duration, however, shows a slight growth from 52% in 7-year-olds to around 64% in ages 10 and above. An anomalous dip occurs at age 12 where durational accuracy drops significantly to 52%.

4.5 Pitch contour as a discriminative cue

Figures 11 and 12 address whether pitch contour alignment functions as a discriminative cue. In the duration panel of figure 11, individual children vary substantially in slope and both group smooths rise clearly. In the pitch panel, nearly all individual curves are horizontal and below 50%. The Step x Trial Type interaction was significant (OR = 0.901, z = -2.63, p = .008). The slope contrast (duration minus pitch = 0.151, z = 6.00, p < .0001) is the key result for RQ3: the pitch-alignment manipulation produces no reliable perceptual change, while the duration manipulation does. Pitch contour, as operationalised through progressive delay of the F0 turning point, is not used as a discriminative cue by these children.

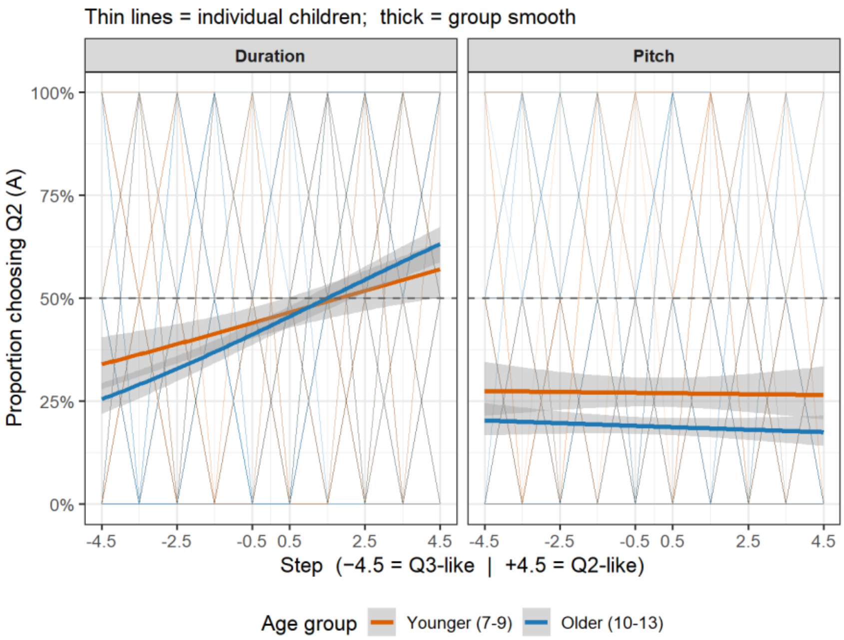

Figure 11. *Individual psychometric curves with group trend.*

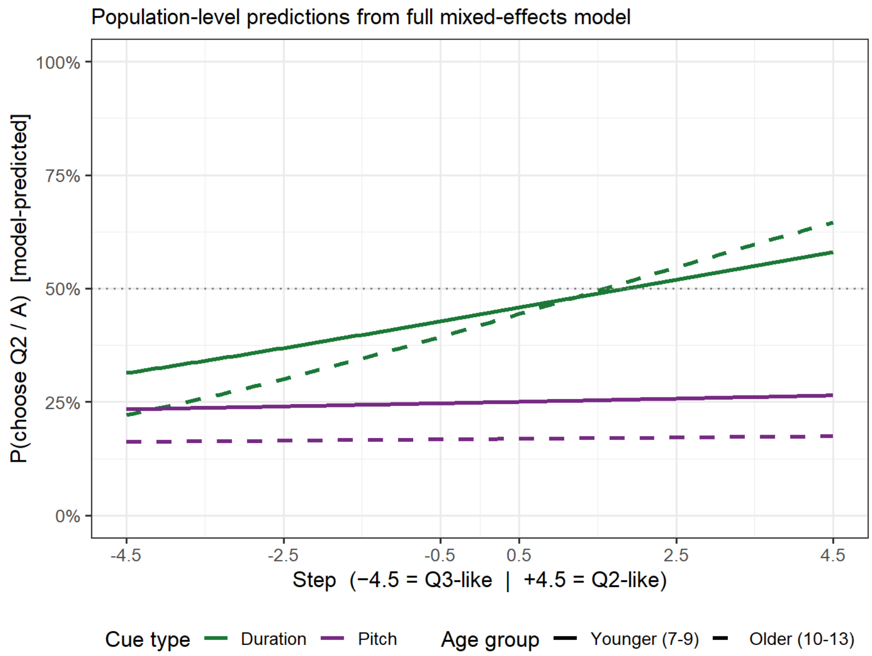

Figure 12. *Model-predicted psychometric curves (population-level predictions from full mixed-effects model).*

Figure 13 shows the forest plot of fixed-effects odds ratios. The pitch vs. duration intercept difference (OR ~ 0.42) reflects the strong global Q3 bias in the pitch condition. The positive step effect (OR ~ 1.13) reflects the duration slope. The Trial Type x Older age interaction (OR ~ 0.67) captures the greater pitch Q3-bias in older children.

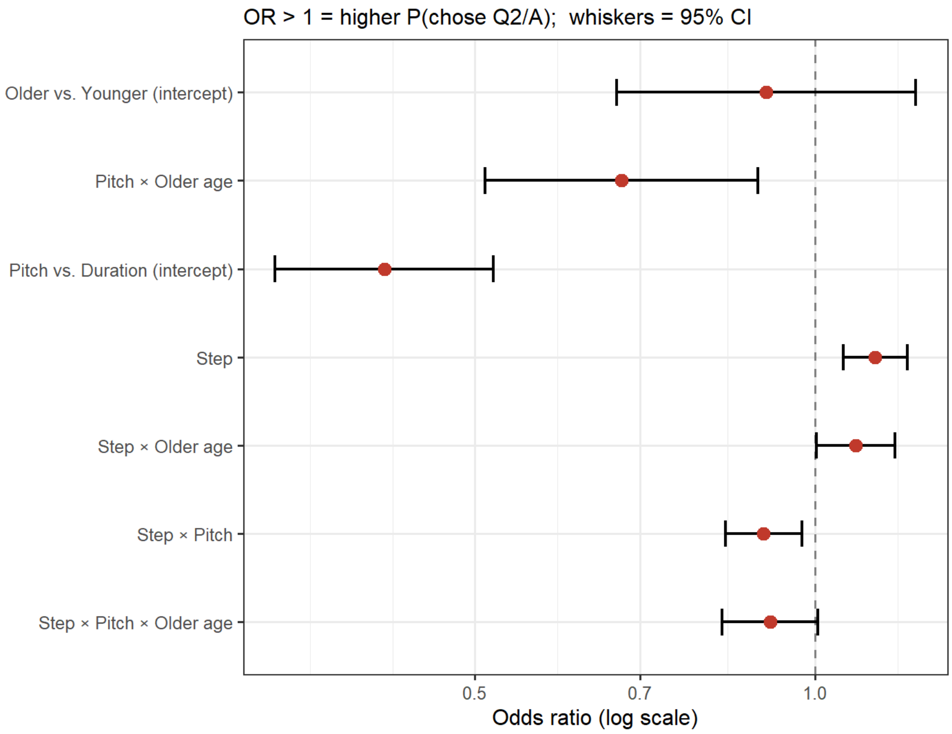

Figure 13. *Forest plot of fixed-effects odds ratios from the full mixed-effects logistic regression.*

4.6 Trends within individual age years

Separate mixed-effects logistic models were fitted for each age year. Table 4 summarizes estimated P(Q2) at the continuum midpoint and duration slope significance for each age year. Because seven individual age models were fitted, the duration p-values reported in Table 4 were corrected for multiple comparisons using the Holm method. Ages 8 to 11 remain significant after correction. The per-age-year results are exploratory.

Table 4. *Per-age-year model results.*

Age	n	P(Q2) Duration	P(Q2) Pitch	Duration slope p	After correction
7	8	0.46	0.31	p = .904	*p = .904*
8	19	0.41	0.18	p = .004**	*p = .020**
9	13	0.48	0.31	p = .005**	*p = .020**
10	30	0.39	0.16	p < .001***	*p = .007**
11	34	0.46	0.14	p < .001***	*p = .007**
12	15	0.41	0.24	p = .126	*p = .342*
13	2	0.40	0.21	p = .114	*p = .342*

Figure 14 presents the mean curves for duration and pitch separately per individual age. The figure reveals a developmental pattern in the duration condition – positive slope being absent at age 7, emerging weakly at age 8 and strengthening progressively through ages 9 and above. By ages 10-11 the durational curve crosses the 50% chance line in the positive step range, indicating above-chance Q2-Q3 discrimination.

The pitch curve, in contrast, remains flat and below 50% at every age, with no meaningful variation in slope across development. Notably, the confidence bands for age 13 are very wide due to extremely small sample (n = 2). These data points should not be interpreted as representative.

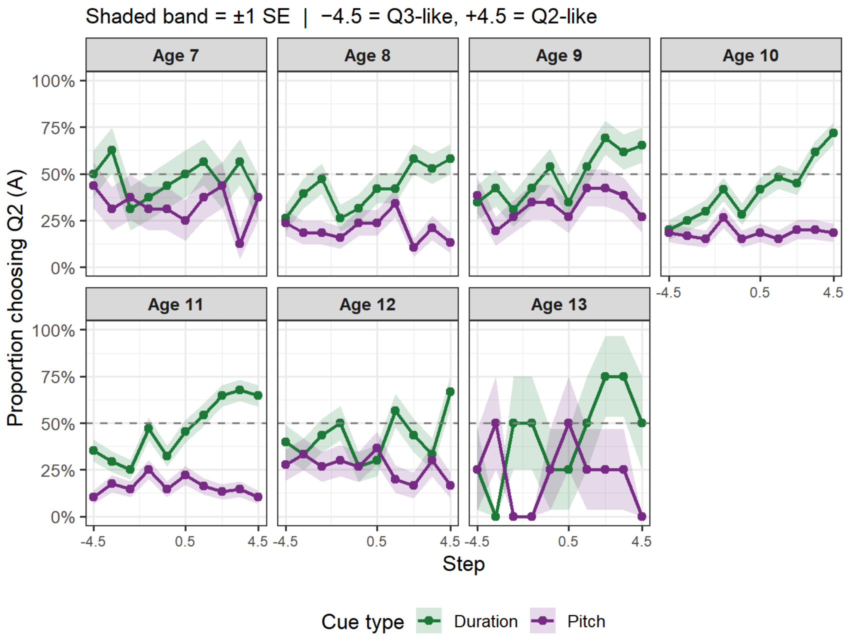

Figure 14. *Psychometric curves by individual age year.*

Figure 15 displays accuracy by age year. For duration, accuracy rises from ~52% at age 7 to a peak of ~64% at ages 10-11, consistent with the pattern in Table 3. A notable dip occurs at age 12, though this is likely a sample size artefact (n = 15), rather than a developmental reversal.

For pitch, accuracy remains near the 50% chance line throughout with no discernible developmental trend.

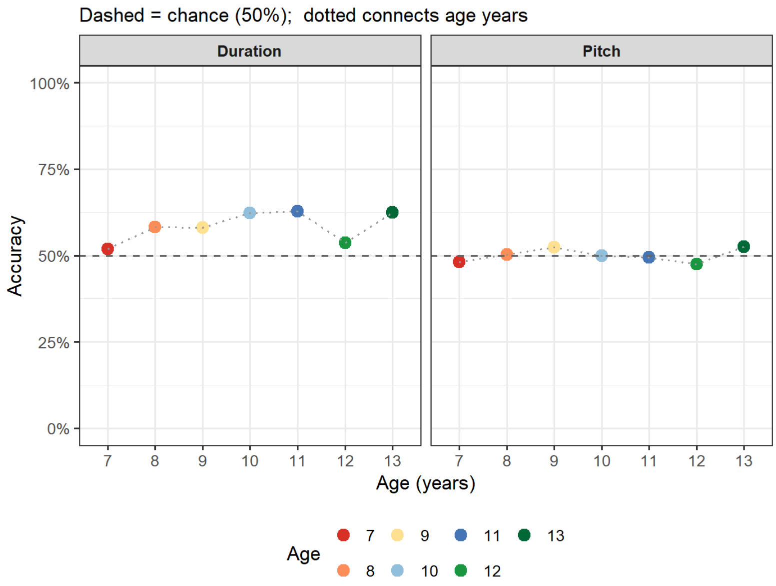

Figure 15. *Accuracy by age year for Duration and Pitch.*

5. Discussion

5.1 Overview

This study examined how L1 Estonian-speaking children aged 7-13 perceive the distinction between Q2 and Q3 vowels, using the AXB forced-choice task with two acoustic continua. The durational continuum progressively shortened the stressed vowel /ɑː/ from its Q3-like baseline toward a Q2-like length across ten equal steps, while the pitch continuum progressively delayed the F0 turning point within the stressed syllable from an early-falling Q3-like position towards a late-falling Q2-like position.

The results converge on three main conclusions: (1) children aged 7 to 13 show partial and developing sensitivity to durational cues for Q2/Q3 discrimination, emerging from age 8 onward; (2) formal education is associated with a modestly steeper durational response slope, but paradoxically with a stronger Q3 bias in the pitch domain; and (3) pitch contour alignment (the cue that is decisive for many adult native listeners) is not used discriminatively by children across the entire age range studied.

5.2 Duration-based perception

Duration psychometric function rises positively and significantly from approximately 31% Q2 responses at step −4.5 to approximately 64% at step +4.5, crossing the 50% chance line between steps +0.5 and +1.5. This is the expected direction for a listener who is sensitive to the manipulation: as the stimulus X becomes progressively shorter and more Q2-like, children increasingly judge it as sounding like A (Q2) rather than B (Q3). This broad positive trend is consistent with the view that durational sensitivity to quantity contrasts develops gradually across the preschool and early school years (Kunnari, Nakai & Vihman 2001; Vihman, Nakai & DePaolis 2006).

However, the function departs from adult-like categorical perception in two important ways. First, it does not achieve the sharp S-shaped characteristic of fully categorical perception. The curve rises gradually and shows irregular noise near the midpoint (a dip around -0.5) indicating that children can detect durational differences but have not yet formed a clearly defined perceptual boundary between Q2 and Q3. Secondly, even among the oldest children, durational accuracy reaches only approximately 64%, substantially below the near-perfect discrimination typical of adult native listeners. This gap suggests that the 7–13 age range represents an extended period of gradual consolidation rather than a completed acquisition.

The within-age-year analysis provides a more detailed developmental picture. The duration slope is non-significant for age 7 ($p = .904$), meaning the youngest children did not show reliable sensitivity to the durational manipulation at all. Significance emerges at age 8 ($p = .004$) and strengthens through ages 9, 10, and 11, before becoming non-significant again at ages 12 and 13. The non-significant slope at age 12 is likely a sample composition effect due to a small sample group size ($n = 15$); the pattern at age 13 ($n = 2$) cannot be interpreted with any confidence. The older group as a whole shows a significantly steeper duration slope than the younger group (Step × Age Group OR = 1.09, $p = .041$), consistent with increased durational sensitivity over the school years, though whether this reflects maturation, cumulative language experience, or the onset of formal instruction - or all three - cannot be established from cross-sectional data alone.

A methodological consideration is that the duration stimuli manipulated only the first syllable vowel, holding the second syllable constant. As discussed in Chapter 3, this simplification is standard in the Estonian quantity perception literature and is justified by evidence that first-syllable acoustic properties are the primary perceptual locus for listeners (Lippus, Pajusalu & Allik 2009, 2011). Nevertheless, the stimuli do not replicate the compensatory duration relationship between the two syllables that characterises natural Estonian speech (Lippus et al. 2013), and it is possible that children's sensitivity to the Q2/Q3 contrast in more naturalistic stimuli is somewhat higher than what the present continuum captures.

5.3 Pitch contour perception

The most striking finding of this study is the near-total absence of pitch-based discrimination. This contrasts with the adult Estonian literature, in which the location of the F0 turning point within the stressed vowel is the primary and sometimes sufficient cue for Q2/Q3 discrimination (Lippus, Pajusalu & Allik 2009, 2011; Eek 1980).

The strong and consistent Q3 bias, present even at the most Q2-like end of the pitch continuum, indicates that children are responding to some feature of the pitch stimuli, but not the feature that varied across steps. The most plausible interpretation is that all stimuli in the pitch continuum retained a falling F0 contour within the stressed vowel to some degree, even those modified toward the Q2-like end. Children appear to associate any falling pitch in the stressed syllable with Q3 as a categorical default, without tracking the fine-grained positional information (early fall vs. late fall) that differentiates Q2 from Q3 in the adult system. This would represent a partially specified phonological representation: children have acquired the association between Q3 and a falling pitch contour but have not yet fully internalized the turning-point location that constitutes the adult perceptual boundary.

This interpretation connects to Lippus' (2011) hypothesis that Estonian is undergoing a typological shift from a quantity language toward a pitch-accent language. If the tonal component of the Q2/Q3 distinction is a relatively recent and still-consolidating development in the ambient language, children in the 7–13 age range may have access to a less fully specified phonological representation of the pitch cue than adult speakers. The present results are consistent with this scenario, though they do not conclusively establish it.

The regional provenance of participants is directly relevant here. Both recruitment locations (Pärnu and Häädemeeste) are in south-western Estonia, the dialect region where adult speakers show the most robust pitch-based Q2/Q3 discrimination (Lippus & Pajusalu 2009; Pajusalu 2009). The fact that even children from this pitch-reliant dialect area show no pitch-based discrimination is therefore particularly intriguing: it suggests that the adult-like pitch-alignment perceptual boundary is not yet established in south-western Estonian children aged 7–13, despite growing up in the dialect community where pitch is the dominant adult cue.

5.4 The role of formal education

The comparison between younger (7–9, pre-instruction) and older (10–13, post-instruction) children reveals a nuanced picture of the role of formal instruction. The significantly steeper duration slope in older children ($p = .041$) is consistent with the hypothesis that schooling sharpens sensitivity to durational cues, though maturation and increased language exposure confound any causal interpretation in a cross-sectional design.

More unexpectedly, older children show a significantly stronger Q3 bias in the Pitch condition than younger children (P(Q2) = 0.168 vs. 0.248, $p = .003$). Several interpretations are possible. First, formal instruction may emphasize the perceptual salience of Q3 in a way that strengthens a global categorical association between falling pitch and Q3 without simultaneously developing sensitivity to the specific turning-point location that distinguishes Q2 from Q3. Estonian school materials describe Q3 as having a distinctive falling pitch contour; children who have internalized this description may apply an 'any falling pitch = Q3' rule more consistently than younger children who lack the metalinguistic category entirely.

Secondly, following Bialystok (2001), the metalinguistic shift demanded by explicit instruction may not straightforwardly improve perceptual sensitivity. The transition from implicit phonological knowledge to explicit metalinguistic categorization is cognitively demanding and may initially produce a more categorical but less finely tuned Q3 response. Thirdly, it is possible that the older group's stronger Q3 bias reflects genuine phonological development: a more consolidated Q3 pitch representation may initially manifest as a bias before a fully specified Q2/Q3 boundary eventually emerges above the age range studied here.

Regardless of the mechanism, the current findings suggest that formal instruction as currently delivered does not produce improved categorical pitch-based

Q2/Q3 discrimination in the 7-13 age range. This points to a possible mismatch between the level at which quantities are taught (primarily as an orthographic-metalinguistic category within literacy instruction) and the level at which Q2/Q3 discrimination would need to operate to improve pitch-based perception, namely the fine-grained acoustic-phonological level of F0 alignment.

5.5 The AXB vs. BXA divergence

The comparison of AXB and BXA response curves (Figure 7 in the Results) revealed a systematic divergence in the duration condition: AXB responses were consistently higher than BXA responses at all steps, while pitch condition curves overlapped closely.

In the AXB order, children heard the unmodified Q2 stimulus first, which may have primed them to categorise the middle stimulus as Q2-like when in doubt. In the BXA order, hearing Q3 first may have had the opposite effect. Such presentation-order priming is well-documented in AXB paradigms and is addressed in the current design by counterbalancing AXB and BXA across participants. The absence of a similar divergence in the pitch condition is consistent with the Q3 bias being so dominant that there is little room for ordering effects to influence participants' responses.

5.6 Limitations and future research

Several limitations of the present study should be noted. First, the sample sizes at the age extremes are small: only 8 seven-year-olds and 2 thirteen-year-olds contributed data, and results for these ages should be treated with caution. Future research should aim for more balanced and larger groups across the age range.

Secondly, the cross-sectional design does not allow for causal claims about whether observed age differences reflect maturation, formal instruction, or accumulated language experience. A longitudinal study following the same children across the transition from pre-instruction (ages 7–9) to post-instruction (ages 10 and above) would provide more direct evidence for the role of formal education.

Thirdly, recruitment was confined to two locations in southwestern Estonia (Pärnu and Häädemeeste). Given previously documented regional variation in cue weighting for Q2/Q3 discrimination (Lippus & Pajusalu 2009), the developmental trajectory of pitch integration may differ in eastern Estonia, where adult speakers rely more heavily on duration. A cross-dialectal comparison would substantially enrich the developmental picture.

Finally, due to the scope of this study, the stimuli manipulated only the first syllable, while the second syllable was held constant. Future research using fully resynthesised foot-level stimuli that vary both syllable durations would allow a more valid assessment of children’s sensitivity to the full durational signature of Q2 versus Q3.

5.6 Conclusion

This study provides an investigation of Q2/Q3 vowel quantity perception across the 7–13 age range in L1 Estonian-speaking children. Durational sensitivity to the contrast develops gradually from age 8 onward, reaching above-chance but not adult-like levels by ages 10–11. Pitch-based discrimination is absent throughout the entire age range. Children treat all pitch-manipulated stimuli as categorically Q3, suggesting that the fine-grained phonological boundary between Q2 and Q3 has not yet been acquired by age 13. Formal instruction is associated with modestly improved durational sensitivity and a paradoxically stronger pitch-based Q3 bias, indicating that the knowledge introduced in school does not straightforwardly translate into improved pitch discrimination.

These findings contribute to the theoretical discussion about Estonian's status as a language in transition between a quantity and a pitch-accent typology (Lippus 2011). If adult-like pitch-based Q2/Q3 discrimination is not yet acquired by age 13, the tonal component of the Estonian quantity system may be genuinely less robustly represented in younger speakers' phonological grammars - which is consistent with the hypothesis that this feature is a recent and still-consolidating development in the prosodic system.

Future research should extend the age range and compare dialect regions to build a more accurate picture of how Estonian-speaking children acquire one of the most complex prosodic systems in the world.

References

Argus, R. (2008). *Eesti keele muutemorfoloogia omandamine* [Doctoral dissertation, Tallinn University]. Tallinn University Press.

Asu, E. L., & Teras, P. (2009). Estonian. *Journal of the International Phonetic Association*, *39*(3), 367–372. https://doi.org/10.1017/S002510030999017X

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Boersma, P., Weenink, D., & Shchupak, A. (2025). Praat: doing phonetics by computer [Computer program].

Eek, A. (1974). Observations on the duration of some Estonian vowel sequences. *Preprints of the Symposium on the Acoustics of Speech*, Tallinn, 1–12.

Eek, A. (1980). Quantity in Estonian: Notes on the perception of duration. *Estonian Papers in Phonetics*, 1, 5–30.

Eek, A., & Meister, E. (1998). Quality of Standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. *Linguistica Uralica*, *34*(3), 226–233.

Kask, L., Põldver, N., Lippus, P., & Kreegipuu, K. (2021). Perceptual asymmetries and auditory processing of Estonian quantities. *Frontiers in Human Neuroscience*, *15*, Article 612617. https://doi.org/10.3389/fnhum.2021.612617

Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th*

Annual Boston University Conference on Language Development (Vol. 1, pp. 412–422). Cascadilla Press.

Lahtein, M., Mailend, M.-L., & Padrik, M. (2021). Kuidas analüüsida eesti laste hääldusepuudeid? [How to analyse speech sound disorders in Estonian children?]. *Eesti Rakenduslingvistika Ühingu aastaraamat*, *17*, 155–175. https://doi.org/10.5128/ERYa17.10

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. In T. A. Sebeok (Ed.), *American studies in Uralic linguistics* (Uralic and Altaic Series, Vol. 1, pp. 21–82). Indiana University Publications.

Lindström, L., & Pajusalu, K. (2003). *Corpus of Estonian Dialects and the Estonian Vowel System*. *Linguistica Uralica, 39(4), 241–257*.

Lippus, P. (2011). *The acoustic and perceptual description of the Estonian quantity system* [Doctoral dissertation, University of Tartu]. University of Tartu Press.

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, *41*(1), 17–28. https://doi.org/10.1016/j.wocn.2012.10.005

Lippus, P., & Pajusalu, K. (2009). Regional variation in the perception of Estonian quantity. In M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Nordic prosody: Proceedings of the Xth Conference, Helsinki 2008* (pp. 151–157). Peter Lang.

Lippus, P., Pajusalu, K., & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, *37*(4), 388–396. https://doi.org/10.1016/j.wocn.2009.07.002

Lippus, P., Pajusalu, K., & Allik, J. (2011). The role of pitch cue in the perception of the Estonian long quantity. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (Studies in Natural Language and Linguistic Theory, Vol. 82, pp. 209–224). Springer. https://doi.org/10.1007/978-94-007-0137-3_10

Lippus, P., Urm, A., Tuisk, T., & Tammiste, M. (2019). Exploring pitch and length perception: An EEG study on quantity discrimination in preschoolers with developmental language disorder. *Brain and Language*, *196*, Article 104652. https://doi.org/10.1016/j.bandl.2019.104652

Meister, L., & Meister, E. (2013). Production of Estonian quantity contrasts by native speakers of Finnish. In *Proceedings of Interspeech 2013* (pp. 2408–2412). ISCA. https://www.isca-archive.org/interspeech_2013/meister13_interspeech.pdf

Pajusalu, K. (2009). Dynamics of Estonian phonology. *STUF – Language Typology and Universals*, *62*(1/2), 109–121. https://doi.org/10.1524/stuf.2009.0008

Pajusalu, K. (2012). Phonological innovations of the Southern Finnic languages. In R. Grünthal & P. Kallio (Eds.), *A linguistic map of prehistoric Northern Europe* (Mémoires de la Société Finno-Ougrienne, Vol. 266, pp. 201–224). Finno-Ugrian Society.

Prillop, K. (2013). Feet, syllables, moras and the Estonian quantity system. *Linguistica Uralica*, *49*(3), 161–186.

Riigi Teataja. (2002). *Eesti keele ainekava 1.–3. klassile*. https://www.riigiteataja.ee/akt/12888846

Soodla, P., Lerkkanen, M.-K., Niemi, P., Kikas, E., Silinskas, G., & Nurmi, J.-E. (2015). Does early reading instruction promote the rate of acquisition? A comparison of two transparent orthographies. *Learning and Instruction*, *38*, 14–23. https://doi.org/10.1016/j.learninstruc.2015.03.001

Themas, L., Schaadt, G., Lippus, P., Padrik, M., Kask, L., Vaher, U., Männamaa, M., & Kreegipuu, K. (2025). *Auditory mismatch response to pitch and duration changes in children with developmental language disorder: A longitudinal approach* [Preprint]. Authorea. https://doi.org/10.22541/au.175367725.54753416/v1

Vihman, M. M., & Kunnari, S. (2011). The sources of phonological knowledge: A cross-linguistic perspective. *Recherches Linguistiques de Vincennes*, *40*, 103–129.

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory phonology 8: Varieties of phonological competence* (pp. 341–366). Mouton de Gruyter.

Winter, M. (2022). *Estonian language*. EBSCO Research Starters. https://www.ebsco.com/research-starters/language-and-linguistics/estonian-language