*Manuscript
Click here to view linked References

Complex prosodic focus marking in Finnish: Expanding the data landscape

Anja Arnhold[1]

Department of Linguistics, University of Alberta, Edmonton T6G 2E7, Canada

Department of Linguistics, Goethe-University, Norbert-Wollheim-Platz 1, 60629 Frankfurt am Main, Germany

Institute of Behavioural Sciences, Speech Sciences, PO Box 9, 00014 University of Helsinki, Finland

Abstract

By investigating prosody beyond pitch and duration, this article provides a detailed and multifaceted picture of focus marking in a language that differs substantially from more extensively studied languages like English. A production study examined prosodic focus marking in Finnish based on acoustic analyses of 947 short SVO sentences spoken by 17 native speakers. The results indicated effects of information structure on five acoustic measures: f0, duration, intensity, the use of pauses and non-modal voice quality. Words in narrow focus had a larger f0 range, longer duration, larger intensity range and were followed by pauses more often than words in other information structural conditions. Conversely, contextually given words showed a smaller f0 range, shorter duration, and, in post-focal condition, lower intensity. Moreover, realisations with non-modal voice quality were more frequent for all syllables of post-focal given words compared to the broad focus condition, whereas for words in narrow focus, non-modal realisations were more frequent only on the last syllable. Observing these effects in parallel, the findings exceeded previous studies in scope, providing encouragement for a broader approach to the investigation of prosodic focus marking.

[1]Tel: +1 780 492 5174; fax: +1 780 492 0806
Email address: anja.arnhold@gmail.com

Keywords: prosody, information structure, f0, duration, voice quality, pauses, intensity

1. Introduction

For communicating successfully, it is vital that speakers take their addressees' knowledge state into account and tailor their utterance to fit the given context. In many languages, prosodic characteristics like pitch or duration signal how an utterance relates to the discourse context, e.g. differentiating new information from information that is already accepted in the common ground shared between discourse participants. For example, a speaker of English will realise a sentence like *I went to visit Mary yesterday* with a different prosody when it answers the question *What's new?* than when answering *Who did you visit yesterday?*. In the second context, the word *Mary* will be in narrow focus, receiving a prominent pitch accent. It will be realised with a wider fundamental frequency (f0) range, longer duration and higher peak intensity than when the sentence answers the first question (broad focus context). At the same time, the other words in the sentence, which are given in the context, will have a reduced f0 range and intensity, and shorter durations. While the propositional content is exactly the same in both instances, the sentence has a different information structure, i.e. it differs in the way the information it conveys relates to the context and the common ground of information shared by the interlocutors.

The present production study investigated prosodic marking of information structure in Finnish. Analysing five acoustic parameters at the same time–f0, duration, intensity, pauses and voice quality—it found effects on all of them: Focused constituents had larger f0 ranges, longer durations, larger intensity ranges, were followed by pauses more often and ended in non-modal voice quality more often. These results add to what is known about prosody and information structure in Finnish. Also from a typological point of view, providing a more detailed picture of this Uralic language expands the data landscape, since information structure marking and prosody are both most extensively studied for

English and other Germanic languages. Methodologically, in finding effects on five acoustic parameters within one carefully controlled experiment, the study encourages a broad approach to studies of prosody which standardly takes several acoustic measures into account at the same time. The present study thus adds to our knowledge of Finnish prosody specifically, but also contributes to extending the scope of research on prosody and information structure from a typological as well as methodological perspective.

Before turning to the study itself, the rest of this section provides the background and motivation for this study. Section 1.1 discusses some basic information structural concepts, as well as a few typological observations about information structure marking, while section 1.2 reviews methodological issues in studying prosodic focus marking. Section 1.3 reviews basic properties of Finnish prosody, as well as the existing literature on prosodic information structure marking in Finnish, based on which section 1.4 derives hypotheses for the present study.

1.1. Information structure

Languages employ different linguistic means to shape utterances in a way that facilitates transmission of the propositional content given the discourse context (called 'information packaging' by Chafe, 1976). In particular, they frequently mark information structure, i.e. the internal organisation of the utterance with respect to the changing information mutually known to be shared by the discourse participants ('common ground', see Stalnaker, 2002, for a review). Several information structural divisions have been proposed (see Krifka, 2008, for discussion). This article concentrates on the notions of focus and givenness. Focus is a part of an utterance indicating that alternatives are relevant for interpretation (Rooth, 1985, 1992; Krifka, 2008). For example, when *yesterday* is focused in *I went to visit Mary yesterday*, relevant alternatives include *I went to visit Mary last week* or *I went to visit Mary ten years ago* (but not *I went to visit Kim yesterday*). For the materials of this study, the focus of a sentence will be identified as the part that is specifically asked for in a question or correcting it, e.g. *yesterday* in *When did you go to visit Mary?—I went to visit Mary yesterday* and in *Did you go to visit Mary last month?—I went to visit Mary yesterday.* These two cases will be referred to as 'information focus' and 'contrastive focus' below (see, e.g. Krifka, 2008). Given information is that which is already part of the common ground shared by the discourse participants, and givenness is in fact frequently viewed as a gradient notion (Gundel et al., 1993; Krifka, 2008, but see e.g. Schwarzschild, 1999 for a non-gradient definition). Here, it suffices to say that when a speaker repeats information from a question in their answer, this material is given. In the example above, and in the materials used here, every part of the utterance that is not focused is given, but the two notions are independent in principle (see e.g. Beaver et al., 2007; Féry & Ishihara, 2009, on second-occurrence focus).

The effects of focus and other aspects of information structure on prosody are most extensively studied for Indo-European and especially Germanic languages like English. In Germanic languages, focus and givenness influence prosody—focused constituents becoming more prominent and given ones less so, as described for the example above—and any description of sentence-level phenomena like accent placement will need to take information structure into account (e.g. Gussenhoven, 1983, for an early article making this case). However, prosodic focus marking is neither universally the same nor even universally present. For example, prosodic focus marking is absent in Northern Sotho, a Bantu language (Zerbian, 2006), and focus marking is generally not obligatory in Hausa, an Afro-Asian language (Hartmann & Zimmermann, 2007). Moreover, while marking of information structure is an important function of prosody in many languages, they employ different strategies like manipulating the presence and types of pitch accents or changing prosodic phrasing (see the typological overviews in Gussenhoven, 2004; Féry, 2013; Jun, 2014). Thus, to understand of the relationship between information structure and prosody, it is important to gather data and in-depth analyses of as many different languages as possible. This article investigates a Uralic language that, as detailed in section 1.3, differs from the well-studied Germanic languages both in terms of its prosodic system and in terms of the means it uses to mark information structure.

1.2. Methodological considerations in the study of prosodic focus marking

Breen et al. (2010) review existing experimental research on prosodic effects of information structure. For production studies, they discuss four main issues. First, studies sometimes do not provide information on acoustic features, but instead only describe phonological annotations, e.g. the occurrence of different accent types. This allows researchers to directly focus on the relevant question of categorically perceived differences. Unfortunately, however, even researchers working in the commonly-adopted framework of Tone and Break Indices (ToBI) annotations may disagree in their usage of categories, for example on when to employ the labels H* vs. L+H* in English (high vs. rising accent, respectively; see Silverman et al., 1992 and subsequent literature for English ToBI; Jun, 2005 for ToBI annotations of several other languages; Pierrehumbert, 1980; Ladd, 2008 for the autosegmental-metrical model of intonation underlying it, which is assumed here as well). As Breen et al. (2010) point out, this makes it difficult to interpret the studies that only refer to ToBI categories without reporting acoustic measurements.

Second, Breen et al. (2010) draw attention to data selection and the amount of data included in the analysis. They criticise experiments recording small numbers of participants, leading to an increased risk of Type II error (their own experiments include data from between 13 and 17 subjects), as well as studies excluding participants based on a priori expectations about their linguistic behaviour with respect to the research question (e.g. sorting out participants with inconsistent accent realisations in studies investigating contrastive accents), which may lead to an increased risk of Type I error.

Third, they suggest that speaker variability should not be dealt with by analysing data from each participant separately. Instead, they recommend using appropriate statistical analyses like linear mixed-effects models (Baayen et al., 2008; Jaeger, 2008). These methods take individual variation into account while testing the significance of effects across this variation, i.e. acoustic manipulations consistently used to mark focus by all participants.

Fourth, Breen et al. (2010) point out that studying only single features of individual words may be problematic, as the production and perception of acoustic prominence may be context-dependent. They state that this may lead to spurious findings as well as a failure to find real differences. Although they do not exemplify these risks, potential scenarios relevant to the present study easily come to mind: For example, a word may carry an f0 peak that is relatively low within the speaker's range, but still clearly higher than the maxima of all other words within the same utterance, and may thus be perceived as prominent. Conversely, a particular utterance may be spoken in a lively manner that is uncharacteristic for the speaker, leading all words to have large f0 ranges. If one considered only measurements from a single word, a real effect might be left uncovered in the first case, whereas a false alarm would occur in the second case, even if general speaker characteristics were taken into account. Regarding the investigation of single acoustic features, it is possible that one acoustic measure by itself does not vary significantly, even though it contributes to an overall significant marking brought about by several measures at the same time. Conversely, a phonetic measure may vary significantly only when participants are prevented from employing other cues, i.e. a trade-off and an additive use of effects are both possible a priori. Again, these effects may be left uncovered when only a single acoustic feature is analysed.

The present experiment avoids these shortcomings: First, it analyses acoustic features directly. Second, it uses data from 17 speakers, excluding only one speaker for a reason unrelated to the objective of the study. Third, it employs linear mixed-effects modelling to account for random variation between participants and items. Fourth, all words in the recorded sentences were analysed and, crucially, several acoustic parameters were analysed for each word. The list of acoustic features considered exceeds both those analysed by Breen et al. (2010), by encompassing voice quality, and those evaluated by existing studies of prosodic focus marking in Finnish, by including the occurrence of pauses. Crucially, all features were analysed for the same data set, so that it can be assessed whether speakers make significant use of several acoustic markers in parallel. Moreover, the present study considered the effects of focus and givenness at the same time, while varying other factors with a known role in Finnish prosody. The study thus exceeds previous studies on Finnish, as well as most studies on English and other Germanic languages, methodologically in several ways.

1.3. Finnish prosody and information structure marking

As is typical for Uralic languages, a prominent feature of Finnish is its binary quantity distinction for both vowels and consonants in most positions (e.g. *muta* [mut̪ɑ] 'mud', *muuta* [muːt̪ɑ] 'another (partitive)', *mutta* [mut̪ːɑ] 'but', *mutaa* [mut̪ɑː] 'mud (partitive)', *muuttaa* [muːt̪ːɑː] 'to change'). But also beyond the lexical level, Finnish prosody differs strongly from English. The present article follows the account suggested in Arnhold (2014), characterising Finnish as a phrase language in the sense of Féry (2010, 2015). That is, it is analysed as language whose intonation is largely determined by invariable tones associated with prosodic phrases. Earlier accounts describe Finnish in terms of pitch accents associated with stressed syllables (which are always word-initial), but crucially, researchers agree that these accents are uniformly of the same type (Välimaa-Blum, 1993; Iivonen, 1998; Suomi, 2005). This contrasts sharply with intonation languages like English, for which descriptions suggest an inventory of contrasting accents expressing different meanings (e.g. Pierrehumbert, 1980 and Grice et al., 2005 propose an inventory with six accents for English and German, respectively; Gussenhoven, 2005 uses eight accents to model Dutch). For example in German, falling f0 on stressed syllables is often said to be used for words in focus, whereas (contrastive) topics are associated with f0 rises on the stressed syllables (Féry, 1993). No distinction of this kind and no accent inventory has been suggested for Finnish.

In addition to its prosodic differences from English, Finnish is also interesting because it possesses several means of marking information structure. Word order is largely determined by information structure, and while the default word order is SVO, all deviations indicate information structural variation (Vilkuna, 1989, 1995). Thereby, non-contrastive foci appear in final position. For example, *Minna näki Matin* 'Minna saw Matti' is a good answer to questions like 'What happened?' (broad focus) or 'Whom did Minna see?' (object focus), but 'Who saw Matti?' is more appropriately answered by *Matin näki Minna* (subject focus). Contrastive foci and other contrastive elements are located in the absolute initial position, preceding other arguments of the verb (e.g. *Minna Matin näki* 'It was Minna who saw Matti', *Matin Minna näki* 'It was Matti that Minna saw' and *Näki Minna Matin* 'Minna did see Matti'). Additionally, some clitics express information structural distinctions; for example, *-pä* marks its host as contrastively focused in *Minäpä siellä kävin* 'It was I who went there' (Nevis, 1986).

Studies have also shown that Finnish prosody is affected by information structural variation. The default f0 contour of a Finnish sentence, appearing in broad focus, is generally characterised as a series of rise-falls, appearing on all content words except the finite verb and decreasing in height over the course of the sentence (Välimaa-Blum, 1993; Iivonen, 1998; Suomi et al., 2008). This regular downtrend is adjusted in utterances with a narrow focus, where the focused word receives a larger f0 range (Välimaa-Blum, 1988, 1993; Vainio & Järvikivi, 2007). Conversely, f0 range is reduced for words preceding and following the focused constituent, with some studies finding that pre- and post-focal rise-falls are always preserved (Mixdorff et al., 2002; Vainio & Järvikivi, 2007), whereas others report flat contours (Suomi et al., 2003; Suomi, 2005) or variation between both cases (Välimaa-Blum, 1988).

Mixdorff et al. (2002), Suomi et al. (2003) and Suomi (2007) additionally report longer durations for words in narrow focus. However, according to Suomi et al. (2008) this use of duration is reserved for contrastive cases, distinguishing contrastive accents from accents realised by f0 only. Furthermore, Vainio & Järvikivi (2007) reported significant effects of focus on intensity. Analysing the difference in peak intensity between the medial and the final constituent in a three-word utterance, they found that compared to the broad focus condition, the fall from the medial to the final constituent was larger when the medial constituent was in narrow focus and smaller when the final constituent was in narrow focus.

Finally, Vainio et al. (2010) analysed the effect of contrastive focus on laryngeal voice quality, estimating glottal excitation by inverse filtering the waveforms of the speech signal. They found less tense and more breathy voice quality for words in narrow focus, while given words had more breathy and less tense voice quality compared to their broad focus realisation.

1.4. Hypotheses

The present study will investigate the effects of narrow focus (information focus and contrastive focus) and of givenness on the acoustic measures duration, f0 and intensity within one experiment. It will additionally analyse the occurrence of pauses and non-modal voice quality for the same materials. Since participants were free to use any and all measures to express the information structural variations, it is possible that significant effects will appear for only some of the measures due to speaker variation or a trade-off between the measures. By contrast, maximally two acoustic measures were investigated simultaneously in existing studies. Assuming, however, that all effects observed in the previous literature can be reproduced, the following hypotheses can be generated:

1. **Duration**: Durations are longer for words in narrow focus and shorter for given words compared to broad focus realisations (Mixdorff et al., 2002; Suomi et al., 2003; Suomi, 2007). Possibly, the narrow focus effect only appears for contrastive focus (Suomi et al., 2003, 2008).
2. **F0**: F0 range is larger for words in narrow focus and smaller for given words compared to broad focus realisations (Välimaa-Blum, 1988, 1993; Suomi et al., 2003; Vainio & Järvikivi, 2007).
3. **Intensity**: Intensity is larger for words in narrow focus and smaller for given words compared to broad focus realisations (Vainio & Järvikivi, 2007).

4. **Voice quality**: Voice quality is affected by focus (Vainio et al., 2010).

The hypothesis regarding f0 can be made more concrete with the assumption that the range expansion in narrow focus affects tonal targets. In accordance with the autosegmental-metrical theory of intonation (Pierrehumbert, 1980), this article assumes that continuous f0 contours can be described as realisations of a series of tonal targets, which are either high (H) or low (L) (see Ladd, 2008, for an introduction). Different tonal specifications of the Finnish rise-falls have been put forward within this theoretical framework: This paper follows Arnhold (2014) in modelling them as realisations of a high and a following low tonal target associated with the prosodic phrase ($H_\phi L_\phi$), leading to hypothesis 2a. Välimaa-Blum (1988, 1993) suggests they are due to rising accents (L+H* or rarely L+H*), so that an f0 minimum and a following maximum are expected (2b). By contrast, Suomi et al. (2008) state that accents include three tonal targets, LHL, i.e. a maximum with a preceding and a following minimum, generating hypothesis 2c. Moreover, Vainio & Järvikivi (2006, 2007) observed that the rising part of the rise-fall was crucial to narrow focus marking in sentence-medial words, while the fall appeared central for final words. They do not discuss phonological implications and the question of tonal targets, but the observed pattern could be assumed to reproduce in the present study, leading to hypothesis 2d.

2a. F0 expansion in narrow focus affects the f0 maximum of a word and the following minimum, leading to larger falls (in line with $H_\phi L_\phi$, Arnhold, 2014).

2b. F0 expansion in narrow focus affects the f0 maximum of a word and the preceding minimum, leading to larger rises (in line with L+H*, Välimaa-Blum, 1988, 1993).

2c. F0 expansion in narrow focus affects the f0 maximum of a word and the preceding as well as the following minimum, leading to both larger rises and larger falls (in line with LHL, Suomi et al., 2008).

2d. In sentence-medial position, the expansion affects the rising part of the rise-falls more (leading to higher f0 maxima and lower preceding minima), while in final position, a larger effect appears for the falls (leading to higher maxima and lower following minima, Vainio & Järvikivi, 2006, 2007).

For voice quality, it is difficult to derive more concrete predictions, since Vainio et al. (2010) used a method of analysis differing from the one employed here. Moreover, note that Epstein's (2002) and Ní Chasaide et al.'s (2011) inverse filtering studies on English observed the opposite of Vainio et al.'s, (2010) findings, i.e. a more tense voice quality in prominent and narrow focus words. Considering voice quality in studies of focus marking is still a very new trend and for most languages, the topic has not been investigated at all.

Pauses have not been studied in the existing literature on prosodic marking of information structure in Finnish. In several other languages more frequent pauses and pauses with longer duration have been observed preceding new and focused constituents (see Romøren & Chen, 2015, for an overview). However, focused constituents are preferably located at the left edge of prosodic phrases in some languages, while the right edge is preferred in other languages (see Féry, 2013, for a general account), so that both pauses preceding words in narrow focus and pauses following them could be expected. The absence of an effect of information structure on the use of pauses is also possible.

In addition to varying information structure, the present experiment also manipulated vowel quantity and analysed words in different positions in the sentence to control for the effects of these two factors. Long vowel quantity was expected to lead to longer vowel and thus word durations, while f0 and intensity downtrends were expected to lead to lower values for words later in the utterance. Moreover, words in final position were expected to have longer durations in accordance with studies describing final lengthening in Finnish (Lehtonen, 1974; Myers & Hansen, 2007; Nakai et al., 2009, 2012), as well as being realised with non-modal voice quality more frequently (creaky, breathy or voiceless realisations, see Lehtonen, 1970; Iivonen, 1998; Myers & Hansen, 2007; Nakai et al., 2009).

2. Methods

This paper reports a production study eliciting different information structural conditions for short sentences with unmarked word order. Participants uttered simple SVO sentences as answers in question-answer pairs eliciting systematic variation in information structure. Duration, f0, intensity, the use of pauses and voice quality were analysed for all words in these sentences.

2.1. Participants

The present results are based on the data of 17 native speakers of Finnish (11 female), aged 20 to 35 years (mean age: 25.41). All participants had grown up in Helsinki or the surrounding area and were university students or recent graduates. None of them was familiar with prosody research. Participants were reimbursed with a gift certificate for cinema tickets.

Two of the 17 speakers identified as bilingual with Finnish being their stronger language, the other languages being Russian and Bulgarian, respectively. Their data was included in the analysis, since their speech did not noticeably differ from that produced by the other participants. Most participants had experience with several languages; note also that children are required to attend both English and Swedish language classes in Finnish schools.

Data from one additional speaker was excluded from the analysis because she seemed uncomfortable with the recording situation and produced prosodic patterns differing strongly from the other participants. Her intonation was generally flat, but included frequent prominent final rises. These rises have the function of projecting further speech and eliciting a response signal from a speaker's conversation partner (Ogden & Routarinne, 2005), corroborating the impression that this participant was seeking feedback on whether she performed the task correctly.

This research was conducted in accordance with the ethical principles of the University of Helsinki.

2.2. Materials

Materials consisted of eight simple SVO sentences (see Table 1). These sentences contained exclusively disyllabic words, with the exception of the four objects *loi.me.a* 'lichen', *lyi.jy.ä* 'lead', *me.nu.a* 'menu' and *le.lu.a* 'toy'.[2] The materials were selected to contain only sonorant segments as far as possible to facilitate analysis of f0, but some of the verbs included obstruents. All words were controlled for vowel quantity in the first syllable, with only long monophthong or diphthong vowels (quantity 2) occurring in sentences 1 to 4 and only short vowels (quantity 1) in sentences 5 to 8. Second syllable vowel quantity could only be controlled for verbs due to morphological requirements. Subjects always appeared in nominative case and included short second syllable vowels. For objects, the partitive marker *-a/ä* was added to the second syllable vowel, which resulted in a long vowel or diphthong for sentences 1, 2, 5 and 8 and formed a syllable of its own for sentences 3, 4, 6 and 7. Thus, the following syllable structures occurred: CVV.CV (subjects), CV.CV (subject and verbs), CVV.CVV (verbs and objects), CV.CVV (objects) and CV.CV.V (objects). The latter two structures were collapsed and treated as disyllabic in the analyses below, see Arnhold (2014) for a statistical comparison finding no differences.

2.3. Procedure

Participants uttered each of the eight sentences as answers to seven different questions, inducing broad focus (all-new condition), narrow information focus on the subject, verb and object, respectively, and narrow corrective focus on the subject, verb and object, respectively (see the examples in Table 2). Stimulus questions manipulated givenness at the same time as focus. Whenever one

[2] Finnish orthography is near-phonemic and will therefore be used here. Long vowels are represented by doubled letters like in *maalaa* [mɑːlɑː] '(s)he paints'. The graphemes *ö*, *ä* and *a* stand for [ø], [æ] and [ɑ], respectively; *v* is usually realised as the approximant [ʋ] and *t* is a dentalised [t̪]. Otherwise the orthography corresponds to IPA notation for the examples given here (see also Iivonen & Harnud, 2005; Suomi et al., 2008).

Table 1: Lexical material used in production experiment 1.

	Subject	Verb	Object	Translation
1	Niilo	maalaa	liinaa	‘Niilo (m) paints (a) cloth.’
	[niː.lo	maː.laː	liː.naː]	
2	Moona	liimaa	naavaa	‘Moona (f) glues lichen.’
	[moː.na	liː.maː	naː.ʋaː]	
3	Maini	neuloi	loimea	‘Maini (f) knitted a blanket.’
	[mai.ni	neu.loi	loi.me.a]	
4	Väinö	jauhoi	lyijyä	‘Väinö (m) ground lead.’
	[ʋæi.nø	jau.hoi	lyi.jy.æ]	
5	Jani	töni	lavaa	‘Jani (m) pushed a platform.’
	[ja.ni	t̪ø.ni	la.ʋaː]	
6	Jimi	luki	menua	‘Jimi (m) read the menu.’
	[ji.mi	lu.ki	me.nu.a]	
7	Manu	hali	lelua	‘Manu (m) hugged a toy.’
	[ma.nu	ha.li	le.lu.a]	
8	Jali	nyki	lanaa	‘Jali (m) tugged a levelling drag.’[a]
	[ja.li	ny.ki	la.naː]	

[a]A levelling drag is a device that is attached to a tractor for flattening a ploughed field. It is also used in road construction for levelling the ground.

constituent was narrowly focused, the other two were mentioned in the preceding question, i.e. given, whereas the constituent in narrow focus was new. A female native speaker of Finnish from Helsinki recorded the stimulus questions.

Table 2: Examples of questions inducing different information structures for the sentence *Maini neoloi loimea* 'Maini knitted a blanket'.

Information structure	Question
Broad focus	Miten iltapäivä meni? 'How did the afternoon go?'
Narrow information focus on the subject	Kuka neuloi loimea? 'Who knitted a blanket?'
Narrow information focus on the object	Mitä Maini neuloi? 'What did Maini knit?'
Narrow information focus on the verb	Mitä Maini teki loimille? 'What did Maini do with the blanket?'
Narrow corrective focus on the subject	Piiako neuloi loimea? 'Did Piia knit a blanket?'
Narrow corrective focus on the object	Neuloiko Maini huivia? 'Did Maini knit a scarf?'
Narrow corrective focus on the verb	Silittikö Maini loimea? 'Did Maini iron a blanket?'

Each participant thus produced $8 \times 7 = 56$ target sentences, interspersed with 110 filler sentences of varying length and complexity. A different pseudo-randomised order was employed for each participant.

At the beginning of each trial, participants silently read a sentence appearing on a computer screen. They then pressed a keyboard button to hear the pre-recorded question through loud-speakers and finally uttered the sentence as an answer to the question. Participants were allowed to listen to the question again and / or to correct themselves before proceeding to the next trial with another button press.

Recordings took place in a recording studio at the department of Musicology of the University of Helsinki in May 2010, using a high-quality headset microphone, placed approximately 7 cm away from the participant's mouth, with a sampling frequency of 44,100 Hz.

2.4. Data editing

First, the recorded sentences were automatically segmented with a forced alignment segmentation process developed at Aalto University (for documentation of the speech recognition programme using this process, see Hirsimäki et al., 2006). Second, these segment boundaries were automatically adjusted and then manually checked based on displays of waveform, spectrogram and auditory criteria using Praat software (Boersma & Weenink, 2010). Word and syllable boundaries were then determined based on these segment boundaries.

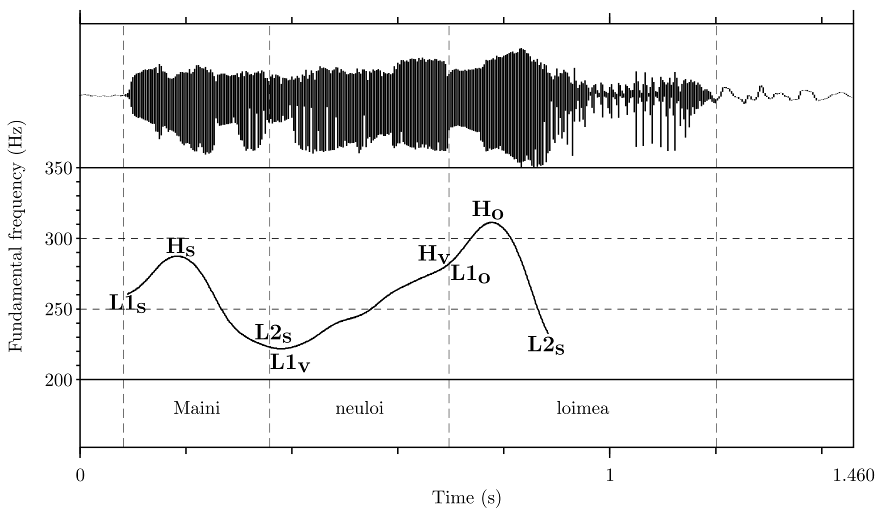


Figure 1: Illustration of f0 measurement points. H = maximum, L1 = preceding minimum, L2 = following minimum of each word. Subscripts mark association with words / constituents: s = subject, v = verb, o = object. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', information focus on the object.

As a basis for f0 analysis, a Praat script marked up to three f0 measurement points for each word: First, the word's f0 maximum (labelled as H), second, the f0 minimum preceding the H point within the same word, if present (labelled as L1), and third, the f0 minimum following the H point within the same word, if present (labelled as L2; see Fig. 1). These three measurement points were chosen to allow for comparison with existing accounts of Finnish intonation, especially Suomi et al.'s (2008) notation of Finnish accents as LHL, i.e. a combination of a low, a high and another low target. This article, following Arnhold (2014), assumes that the H and L2 measurement points generally corresponded to tonal targets associated with the prosodic phrase (denoted as H_ϕ and L_ϕ), while L1 did not correspond to a tonal target in this account.

To eliminate errors produced by Praat's f0 tracking algorithm, the author manually verified the automatic detections for half of the sentences, while a native Finnish speaking student of phonetics with no knowledge of the purpose of the study checked the other half. At the same time, these two annotators marked pauses and stretches of speech uttered with non-modal voice quality (e.g. creaky voice or whisper) on the basis of waveform, spectrogram and auditory impression. Since the three f0 measurement points were all defined objectively as maxima and minima, manual corrections were only necessary in cases of pitch tracking mistakes, e.g. when Praat calculated f0 for voiceless segments or during stretches of non-modal voice quality.

3. Results

The analysis considered altogether $17 \times 56 - 5 = 947$ sentences, with one sentence accidentally being skipped during the recording session and four sentences removed due to hesitations or slips of the tongue (less than 1% lost data). Thus, $3 \times 947 = 2841$ words and $2 \times 2841 = 5682$ syllables were evaluated.

The data was analysed with respect to duration, the use of pauses, f0, intensity and voice quality. These measurements were modelled as dependent variables with linear mixed-effects models implemented in the programme R (Baayen et al., 2008; R Core Team, 2014), as provided by the package lme4 (Bates et al., 2014). These models are extensions of linear regression models.

Like in normal linear regression, they model the observed values with functions containing independent variables (or predictors) as coefficients, as well as a random error term and an (optional) intercept. These parameters are estimated so that the model fits the data optimally, i.e. so that the values predicted from the model have a maximum likelihood to be the values that were actually observed (see Baayen et al., 2008, on maximum likelihood). Thus, the estimation of the parameters models the relationship between the dependent variable and the predictor variables by estimating the size and direction of an adjustment to a baseline value in each condition, i.e. the effect of the manipulation of condition. Mixed-effects models also include random effects variables in addition predictor variables (fixed effects). This allows taking the association between data points coming from the same participant or item into account (Baayen et al., 2008). That means that the model additionally estimates a value by which the estimated effects of the predictor variables are adjusted for every participant or item. Thus, the models test the consistent effects of independent variables across variation between participants and items (e.g. some participant having an overall slower speech rate or higher f0 register than others). Linear mixed-effects models are therefore ideal for analysing experiments with repeated measures from the same participant. The models further allow the analysis of numeric as well as categorical response variables, and can also include both numeric and categorical variables as predictors. Moreover, they give information on size and directionality of effects and also perform well with unbalanced data and missing data points (see Baayen et al., 2008; Jaeger, 2008, on advantages over other analysis methods for linguistic data).

The analyses determined the model with the best fit to the data for each measure. To this end, models with different predictor variables and random effects were computed and compared with likelihood ratio tests carried out by the ANOVA function. These comparisons used log likelihood as a measure of goodness of fit (see e.g. Baayen, 2008). In an implementation of Occam's razor, simpler or parsimonious models were preferred over more complex models containing more variables or interactions, unless adding complexity significantly improved model fit (see Bates et al., ms, for a recent explanation of the importance of parsimonious models). Thus, only variables significantly improving model fit were retained in the best-fitting models reported below. Tested predictor variables always included the three experimental factors, i.e. information structural condition of the word (levels: broad focus, narrow information focus, narrow corrective focus, given while another word was in narrow information focus, given while another word was in narrow corrective focus), first syllable vowel quantity (levels: quantity 1, quantity 2) and position / grammatical role (levels: position 1 / subject, position 2 / verb, position 3 / object). In the analyses of f0 measurements, speaker gender was additionally included as a predictor. Tested random effects were participant, lexical item and trial number (centred to its median). Models with more complex random effects were also computed, adjusting the models for by-participant or by-item differences in the effects of the experimental variables, but since this never affected the overall significance of the factors, the simpler models are reported here. For binomial models of categorical dependent variables, fit with the glmer function, the model output includes p-values based on z values. For the models of numeric dependent variables, fit with the lmer function, p-values calculated based on Satterthwaite's approximation were obtained with the package lmerTest (Kuznetsova et al., 2015). The following significance codes are employed: $*** = p < .001$, $** = p < .01$, $* = p < .05$, $. = p < .1$.

3.1. Duration

Fig. 2 shows mean durations of words and segments in all experimental conditions, with dark grey bars representing consonant durations and light grey bars showing vowel durations. Each panel compares durations in three information structural conditions: given, broad focus (broadF) and narrow focus (narrowF). The two types of narrow focus, corrective and information focus, are collapsed for the sake of simplicity in figures here and below since their effects were very similar. In statistical modelling, the distinction was maintained to assess the similarity of the two narrow focus types, and any differences will be discussed in the text. The section will first analyse word durations and then turn to results for the individual segments.

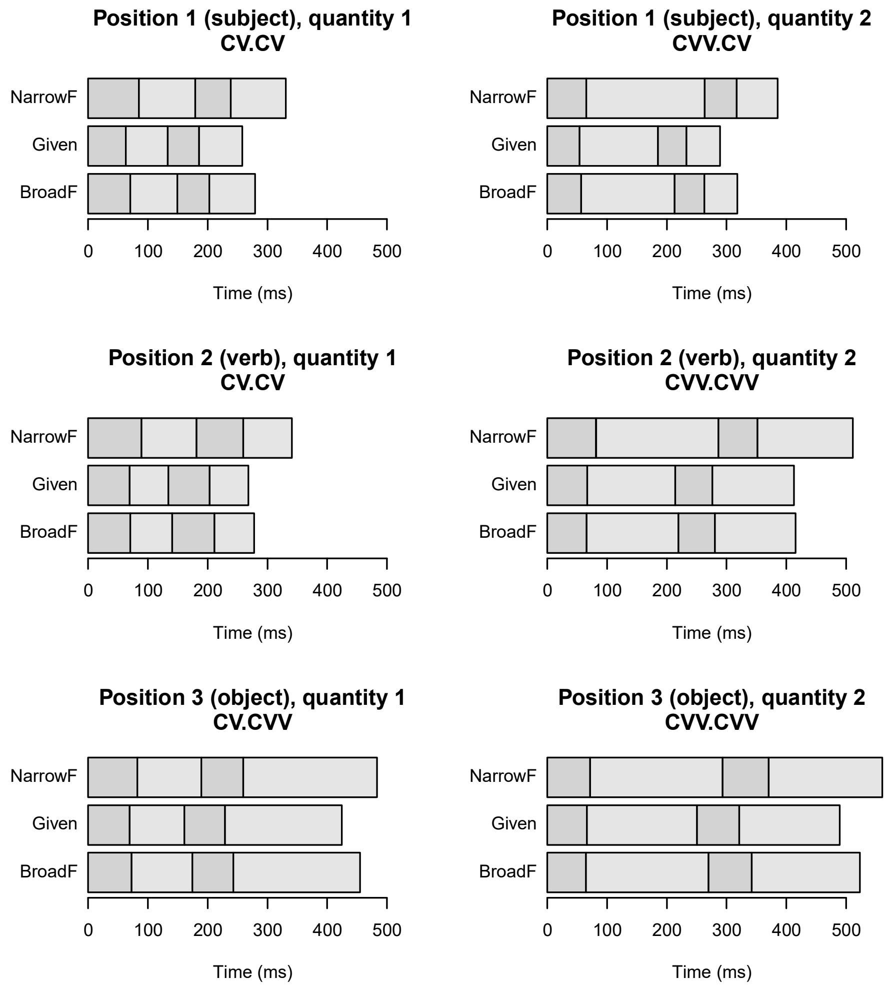


Figure 2: Average segment durations for all experimental conditions.

Table 3 summarises the best linear mixed-effects model of word duration. The second column lists the coefficients for the parameters that the model has estimated. The first value is the intercept, which represents a group mean for the default case when all predictors are at the reference level. Its value of 282 ms corresponds quite well to the mean word duration of 279 ms for words in broad focus in sentence-initial position containing a quantity 1 vowel in their first syllable. The estimates in the following rows give the values by which the intercept is adjusted for the other levels of the predictors. Thus, the coefficient associated with narrow contrastive focus is 56 ms, and indeed sentence-initial quantity 1 words in narrow contrastive focus had a mean duration of 337 ms, i.e. 58 ms more than in broad focus. Narrow information focus is likewise estimated to have a positive effect, while the coefficients of the given conditions are both negative, meaning the intercept is adjusted by subtracting 21 ms and 28 ms, respectively, for observations in these conditions, and so on. The effects of different predictors and their interactions are combined, so that e.g. for sentence-final words in narrow information focus, the intercept is adjusted by $44 + 176 - 29 = 191$ ms. For each observation, the value estimated by the model thus consists of the intercept adjusted by the combined effects of all the predictors, as well as by-participant and by-item adjustments. The third column in Table 3 gives the standard errors associated with each coefficient (with smaller standard errors corresponding to smaller confidence intervals), which are used to calculate the t-values shown in the fourth column. The last column gives p-values and codes whether the coefficient is significantly different from zero. A significant difference indicates that there is a relation between the predictor and the dependent variable (here: word duration) and thus whether the coefficient is a useful predictor (see Baayen et al., 2008; Baayen, 2008, for more details on interpreting summaries of linear mixed-effects models).

The model suggested that all three factors—focus condition, position and quantity—significantly affected word duration, as well as finding interactions between them. Words in both narrow corrective focus and narrow information focus were significantly longer than words in broad focus. Across both quantities and all three positions, the difference was 65 ms on average for corrective focus and 50 ms for information, corresponding to a 17% increase in duration and a 13% increase, respectively. Conversely, the duration of given words was significantly shorter. The average overall difference from broad focus was 21 ms, i.e. a 6% decrease. The differences between focus conditions were notably smallest for

Table 3: Coefficients of the best linear mixed-effects model of word duration (in ms) with random effects of participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	$Pr(>\lvert t \rvert)$	
(Intercept)	281.5451	9.9018	28.4338	0.0000	***
Narrow focus (CF)	55.9333	6.4938	8.6133	0.0000	***
Narrow focus (IF)	43.8826	6.4856	6.7662	0.0000	***
Given (CF)	−21.1807	5.6201	−3.7687	0.0002	***
Given (IF)	−27.8727	5.6225	−4.9573	0.0000	***
Position 2	−7.8253	8.7958	−0.8897	0.3775	
Position 3	175.7679	8.7958	19.9832	0.0000	***
Quantity 2	34.6464	8.2688	4.1900	0.0001	***
Narrow focus (CF) : Position 2	18.2855	7.9644	2.2959	0.0218	*
Narrow focus (IF) : Position 2	21.6317	7.9568	2.7186	0.0066	**
Given (CF) : Position 2	18.3599	6.8971	2.6620	0.0078	**
Given (IF) : Position 2	18.9312	6.8907	2.7473	0.0060	**
Narrow focus (CF) : Position 3	−24.4802	7.9717	−3.0709	0.0022	**
Narrow focus (IF) : Position 3	−28.6221	7.9493	−3.6006	0.0003	***
Given (CF) : Position 3	−11.5682	6.8950	−1.6778	0.0935	.
Given (IF) : Position 3	−2.8898	6.8928	−0.4192	0.6751	
Position 2 : Quantity 2	111.9795	10.0241	11.1710	0.0000	***
Position 3 : Quantity 2	29.5575	10.0241	2.9486	0.0072	**
Narrow focus (CF) : Quantity 2	21.5732	6.5030	3.3174	0.0009	***
Narrow focus (IF) : Quantity 2	16.1018	6.4949	2.4791	0.0132	*
Given (CF) : Quantity 2	−1.9904	5.6318	−0.3534	0.7238	
Given (IF) : Quantity 2	−1.3937	5.6272	−0.2477	0.8044	

sentence-final objects, with narrow focus being 33 ms longer (7% increase) and given realisations 32 ms shorter (7% decrease) than broad focus ones on average, as Fig. 2 illustrates. This is reflected in Table 3 by significant interactions between the factors focus condition and position. The interactions indicated that the effect of both types of narrow focus was indeed significantly smaller for sentence-final objects. For sentence-medial verbs, narrow focus realisations were 80 ms longer than broad focus realisations on average (23% increase), while given ones were only 6 ms or 2% shorter. Accordingly, a further set of interactions between focus condition and position suggested that the effect of both types of narrow focus was significantly larger for verbs than for the sentence-initial subject intercept, whereas the effect of givenness shortening word duration was weaker. Thus, the duration of broad focus verbs was more like that of given ones, while broad focus objects were more similar to narrow focus ones.

Additionally, Fig. 2 illustrates an effect of quantity, primarily affecting vowels: Quantity 2 vowels had longer durations than quantity 1 vowels, as discussed further below. Compared to words with CV.CV syllable structure, mean duration was 35 ms, 160 ms and 192 ms longer for CVV.CV, CV.CVV and CVV.CVV words, respectively (35%, 56% and 67% increase in duration). Testing the effect of the manipulated experimental factor first syllable vowel quantity, the model in Table 3 confirmed that words with quantity 2 vowels in their first syllables had significantly longer durations than those with quantity 1 vowels as their first syllable nuclei. The factor quantity also significantly interacted with focus condition, indicating that narrow focus lengthening had a larger effect in quantity 2 words, i.e. in words with an overall longer duration.

Finally, the main effect of the factor position also appeared as significant, indicating that the duration of sentence-final objects was overall longer than that of sentence-initial subjects. Note that all second syllable vowels were phonologically long in sentence-final position, whereas all sentence-initial subjects included quantity 1 vowels in their second syllables. This connection between quantity and position was reflected in interactions between the two factors (see Table 3). Nevertheless, Fig. 2 illustrates a final lengthening over and above the inherent quantity differences. Thus, CVV.CVV words were 73 ms longer in final position than sentence-medially (17% increase).

Separate analyses of segments indicated that the main effects of information structure on word duration also appeared for the individual segments. Duration of all segments was significantly longer in both narrow focus conditions, again with a larger effect for contrastive focus than for information focus (cf. Tables A.13–A.16 in Appendix A). Shortening in given conditions reached significance for consonants and vowels in the first, but not in the second syllable. In narrow corrective focus, mean segment durations were 14 ms longer for first syllable consonants than in broad focus, 28 ms longer for first syllable vowels, 5 ms longer for second syllable consonants and 18 ms longer for second syllable vowels (corresponding to an increase of 21%, 22%, 8% and 14%, respectively). In narrow information focus, first syllable consonants were on average 11 ms longer, first syllable vowels were 23 ms longer, second syllable consonants were on average 4 ms longer and second syllable vowels were 13 ms longer (16%, 18%, 7% and 10% increase, respectively). In given words, first syllable consonants were on average 2 ms shorter, first syllable vowels were 13 ms shorter, second syllable consonants were on average 1 ms shorter and second syllable vowels were 6 ms shorter (3%, 10%, 1% and 5% decrease, respectively). Thus, information structural effects were largest for first syllable vowels and smallest for the consonants immediately following them.

Final lengthening likewise affected most segments in the word; with the exception of first syllable onset consonants, all segments had significantly longer durations in position 3 compared to position 1. This increase in mean duration was 6 ms for first syllable consonants, 32 ms for first syllable vowels, 19 ms for second syllable consonants and 122 ms for second syllable vowels (10%, 27%, 36% and 177%, respectively). The increase was very large for word-final vowels, where it can be attributed to quantity in addition to position. Compared to second syllable quantity 2 vowels of sentence-medial verbs, those in final position were 33 ms longer (31% increase).

Effects of quantity, as well as interactions, generally showed more variation, as expectable from the make-up of the materials. Recall that while first syllable vowel quantity was manipulated systematically, second syllable quantity was restricted by morphology, so that only quantity 1 vowels appeared in the second syllables of sentence-initial words and only quantity 2 vowels occurred in second syllables of sentence-final words (cf. section 2.2). Therefore, all statistical models tested the effect of first syllable vowel quantity, as this was the independently manipulated factor. Hence, it not surprising that of the models of separate segments, the one of first syllable vowel durations most resembles the model of word duration (compare Tables A.14 and 3). Expectedly, first syllable vowels had significantly longer durations when they were phonologically long. By contrast, first syllable onsets and second syllable vowels were significantly shorter when the first syllable vowel had quantity 2, while second syllable consonants showed no effects. This is in line with previous studies finding that in addition to affecting the duration of the segment itself, vowel quantity is connected to minor durational adjustments of other segments within a longer stretch generally identified with the foot (see Lehtonen, 1970; Suomi, 2009, for details).

Altogether, this section has demonstrated effects of all three factors on word and segment durations: Words in both types of narrow focus and in sentence-final position were significantly longer, whereas given words were significantly shorter. Moreover, the data showed that phonological lengthening of vowels resulted in longer durations of these vowel segments and the words containing them, but also affected other segments in the same word.

3.2. F0 scaling

This section analyses f0 measurements at three points for each word in the data set: the f0 maximum (H), the minimum preceding H within the same word (L1) and the minimum following H within the same word (L2; see Fig. 1 for an illustration). These three measurement points were not always present for each word, e.g. for a word with falling pitch throughout, no L1 value was measured. For some words, no f0 measurements were possible at all, for example due to creaky voice (see section 3.3). Overall, 2081 measurements were obtained for L1, 2533 for H and 2245 for L2. All measured f0 values were converted to semitones (st) relative to 50 Hz for statistical analyses.

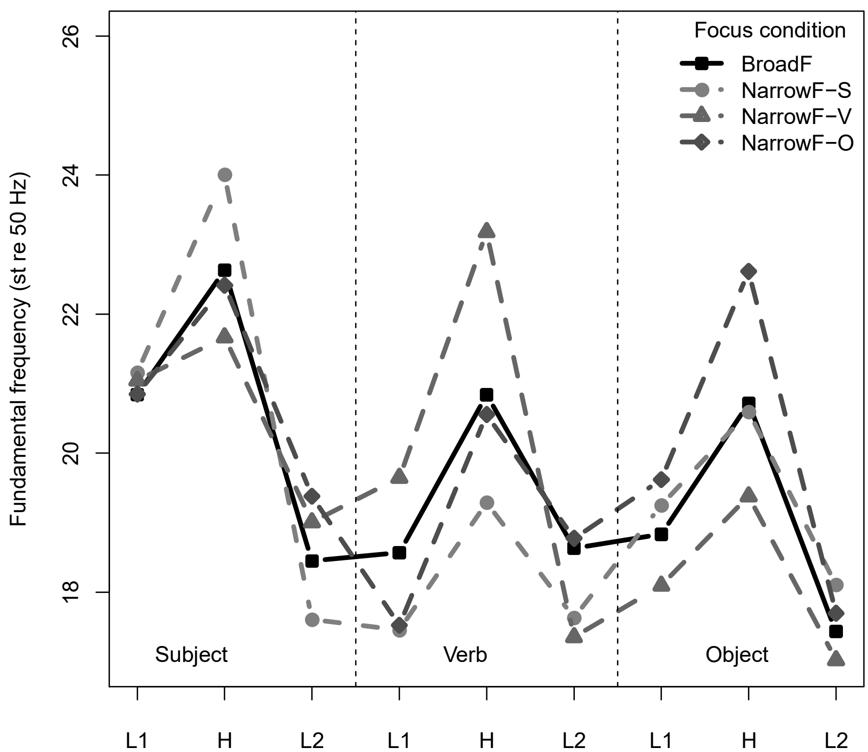


Figure 3: Average f0 for the three measurement points for subjects, verbs and objects in different focus conditions.

Fig. 3 shows the average values for these measurable points for subjects, verbs and objects in broad focus condition and with narrow focus on each of the three constituents (information focus and corrective focus are again collapsed for the sake of simplicity). It illustrates effects of both position and focus condition, confirmed by linear-mixed effect modelling. Most noticeably, for each of the three constituents, the average H value of the narrow focus condition exceeded that of the other conditions by far. Thus, the line representing subject focus in Fig. 3 diverges from the other conditions on the subject. Likewise, verb focus is clearly separate from the other conditions on the verb, while for sentence-final objects, the highest average f0 appears for object focus. In line with this, H measurements were significantly higher for both types of narrow focus than in the broad focus condition, while maxima of given constituents were frequently, but not consistently lower (see Table 4). Additionally, the effects of narrow focus were significantly stronger in second and third position than for the sentence-initial subjects (with the exception of information focus on final objects, where the interaction did not reach significance). H was overall significantly lower in medial and final position than sentence-initially; cf. the three H points for the broad focus condition, represented by the solid black line, in Fig. 3. Also notice the relation between the peaks showing the values for narrow focus on each of the three constituents, respectively, indicating a general f0 downtrend over the course of the sentence. In addition to the effects of focus condition and position, the model of H f0 in Table 4 included a positive main effect of quantity, indicating higher H values for words with quantity 2 vowels, that was however counteracted by several interactions. Lastly, values were overall significantly lower for male speakers.

Whereas H points were higher in narrow focus and (insignificantly) lower in the given condition, converse effects appeared for L2 measurements, i.e. significantly lower values for both types of narrow focus and significantly higher values in given conditions (cf. the model in Table 5). Thus, pitch range was expanded in narrow focus and compressed for given constituents, as illustrated in Fig. 4. In addition to the main effects of focus condition, the model in Table 5 also indicated that L2 measurements were significantly lower for sentence-final objects than for sentence-initial subjects, although no significant difference appeared for the verbs, which also showed a weaker effect of givenness. In sentence-final position, all effects of focus condition were significantly weaker. As Fig. 3 shows, sentence-final minima did not diverge in different conditions, but virtually con-

Table 4: Coefficients of the best linear mixed-effects model of f0 of H (in st). Random effects: participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2533$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	26.9617	0.6483	41.5906	0.0000	$***$
Narrow focus (CF)	1.6949	0.2656	6.3815	0.0000	$***$
Narrow focus (IF)	1.5017	0.2659	5.6470	0.0000	$***$
Given (CF)	−0.3210	0.2327	−1.3795	0.1679	
Given (IF)	0.0972	0.2322	0.4186	0.6756	
Position 2	−1.7664	0.2673	−6.6081	0.0000	$***$
Position 3	−2.0061	0.2693	−7.4503	0.0000	$***$
Quantity 2	0.7339	0.2194	3.3448	0.0010	$**$
Male gender	−11.4313	0.9532	−11.9922	0.0000	$***$
Narrow focus (CF) : Position 2	1.1976	0.3257	3.6771	0.0002	$***$
Narrow focus (IF) : Position 2	0.8404	0.3257	2.5803	0.0099	$**$
Given (CF) : Position 2	−0.4530	0.2867	−1.5801	0.1142	
Given (IF) : Position 2	−0.4976	0.2862	−1.7387	0.0822	.
Narrow focus (CF) : Position 3	0.8725	0.3276	2.6634	0.0078	$**$
Narrow focus (IF) : Position 3	0.5429	0.3270	1.6602	0.0970	.
Given (CF) : Position 3	−0.9631	0.3045	−3.1625	0.0016	$**$
Given (IF) : Position 3	−1.3123	0.2957	−4.4384	0.0000	$***$
Narrow focus (CF) : Quantity 2	−0.6175	0.2669	−2.3138	0.0208	$*$
Narrow focus (IF) : Quantity 2	−0.5119	0.2667	−1.9195	0.0550	.
Given (CF) : Quantity 2	−0.8886	0.2421	−3.6702	0.0002	$***$
Given (IF) : Quantity 2	−0.9380	0.2388	−3.9274	0.0001	$***$

verged on a single level when the f0 fall was cut off by sentence-final non-modal voice quality (see section 3.3). Finally, the model in Table 5 also indicated lower f0 measurements for male speakers and included interactions involving quantity. Quantity 2 generally led to longer durations (see section 3.1), which allowed for longer pitch falls. Thus, the narrow focus effect lowering L2 was larger in quantity 2. However, the effect of quantity was again less strong sentence-finally.

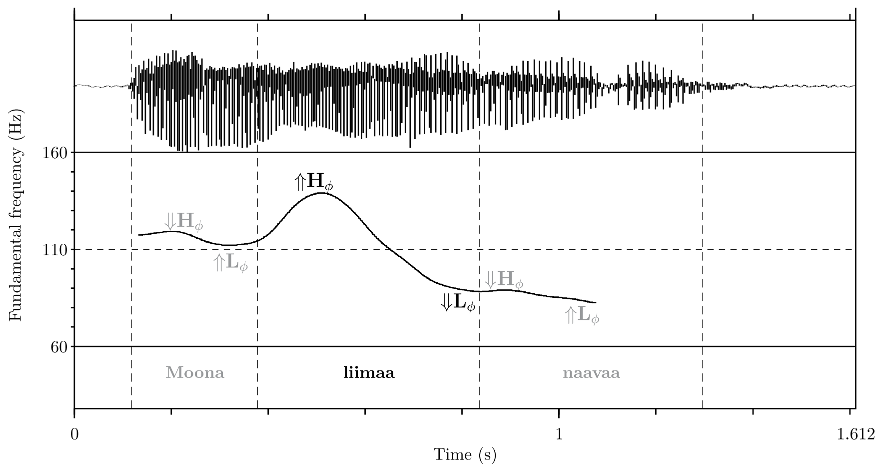


Figure 4: Pitch range expansion on a word in narrow focus (black) and pitch range compression on unfocused words (grey). Sentence *Moona liimaa naavaa* 'Moona glues lichen', narrow contrastive focus on the verb.

As shown in Fig. 4, the present analysis assumes that the measurement points H and L2 generally corresponded to phonological targets, a high and a low tone associated with the prosodic phrase, i.e. $H_\phi L_\phi$ (see Arnhold, 2014, for details). By contrast, it suggests that L1 measurements did not correspond to realisations of targets, but instead were part of interpolations between two targets. In line with this interpretation, mixed-effects modelling indicated that f0 measurements at L1 points were best predicted by the f0 at the two neighbouring measurement points, i.e. L2 of the preceding word and the following H, while adding experimental factors did not improve model fit. Since including a preceding word's L2 as a predictor excluded values from sentence-initial sub-

Table 5: Coefficients of the best linear mixed-effects model of f0 of L2 (in st). Random effects: participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2245$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	23.2564	0.5479	42.4485	0.0000	$***$
Narrow focus (CF)	−0.7520	0.2678	−2.8079	0.0050	$**$
Narrow focus (IF)	−0.6224	0.2709	−2.2973	0.0217	$*$
Given (CF)	0.7165	0.2408	2.9759	0.0030	$**$
Given (IF)	0.9824	0.2391	4.1086	0.0000	$***$
Position 2	0.3289	0.2944	1.1172	0.2660	
Position 3	−1.4950	0.2897	−5.1604	0.0000	$***$
Quantity 2	−0.0714	0.2635	−0.2708	0.7872	
Male gender	−11.2966	0.7830	−14.4282	0.0000	$***$
Narrow focus (CF) : Position 2	−0.0934	0.3372	−0.2770	0.7818	
Narrow focus (IF) : Position 2	−0.1871	0.3391	−0.5517	0.5812	
Given (CF) : Position 2	−1.4058	0.3053	−4.6050	0.0000	$***$
Given (IF) : Position 2	−1.1973	0.3057	−3.9163	0.0001	$***$
Narrow focus (CF) : Position 3	1.5823	0.3293	4.8046	0.0000	$***$
Narrow focus (IF) : Position 3	1.4358	0.3297	4.3550	0.0000	$***$
Given (CF) : Position 3	−1.2270	0.3277	−3.7444	0.0002	$***$
Given (IF) : Position 3	−0.8908	0.3097	−2.8764	0.0041	$**$
Position 2 : Quantity 2	−0.4756	0.2655	−1.7915	0.0872	.
Position 3 : Quantity 2	0.5741	0.2735	2.0992	0.0464	$*$
Narrow focus (CF) : Quantity 2	−1.0743	0.2721	−3.9476	0.0001	$***$
Narrow focus (IF) : Quantity 2	−0.7409	0.2728	−2.7159	0.0067	$**$
Given (CF) : Quantity 2	−0.0338	0.2573	−0.1316	0.8953	
Given (IF) : Quantity 2	−0.4420	0.2526	−1.7494	0.0804	.

Table 6: Coefficients of the best linear mixed-effects model of f0 of L1 (in st). Random effects: participant and item. $N = 1139$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.8456	0.4137	4.4612	0	***
F0 of preceding L2	0.5757	0.0227	25.3561	0	***
F0 of following H	0.3179	0.0166	19.1410	0	***
Male gender	−1.3680	0.2306	−5.9316	0	***

jects, the resulting model accounted for only 1139 of the originally measured 2081 L1 values (cf. Table 6). It did, however, provide an equally good fit as the best model using only experimental factors and gender as predictors, a complex model including several of the effects observed for H, as well as a number of three-way interactions (see Table B.17 in Appendix B).

For the sake of completeness, the models of H and L2 values in Tables 4 and 5 were compared to models including f0 values at neighbouring measurement points as predictors. For L2, ANOVA comparisons indicated no improvement in model fit. For H, including the values of neighbouring L measurements predictors led to a significantly improved fit. While showing a positive correlation between H and the two neighbouring L tones, other effects were largely unchanged in this model (cf. Table B.18 in Appendix B). However, it showed fewer significant interactions between focus condition and position.

In sum, this section has demonstrated effects of the experimental manipulations on f0 scaling, most crucially adjustments of the f0 maximum and the following minimum, leading to a larger pitch range in narrow focus and a reduced range for given constituents.

3.3. Voice quality

Two annotators marked stretches with non-modal (e.g. creaky or breathy) voice quality based on waveform, spectrogram and auditory impression. This section evaluates the effects of the experimental manipulations on voice quality in two ways: First, it analyses the duration of non-modal stretches and second, it examines the location of non-modal stretches within words.

Fig. 5 displays the duration of non-modal stretches in different information structural conditions for the three positions. The top panel shows the absolute duration within the word. The value of this measure was zero for words realised with modal voice quality throughout, while for words with more than one non-modal stretch, it corresponded to the summed duration of non-modal stretches during this word. The bottom panel relates the duration of non-modal stretches to word duration, showing the percentage of word duration realised with non-modal voice quality. Statistical analysis evaluated the simpler measure, i.e. absolute duration in ms, resulting in the linear mixed-effects model in Table 7.

Both panels of Fig. 5 illustrate effects of information structure and position, which were confirmed by the model. Compared to sentence-initial subjects, non-modal stretches were significantly longer for verbs and object, with this effect being larger for the sentence-final objects. For all three positions, differences between information structural conditions are visible in the figure (again, collapsing information focus and corrective focus for the sake of simplicity). Whereas values concentrated around zero for the first position in the sentence, meaning participants largely realised subject constituents without non-modal voice quality, larger values were more frequent when these subjects were in narrow focus. For verbs, longer non-modal stretches appeared in verb focus and subject focus condition, while the object focus condition patterned with broad focus. Objects overall had the longest non-modal stretches and showed the largest values in subject and verb focus conditions. These patterns suggest a distinction between pre-focal and post-focal givenness: Participants realised longer stretches of given words with non-modal voice quality in post-focal positions, but not pre-focally. The difference between verbs in subject-focus and object-focus conditions is particularly revealing, but this generalisation also accounts for the overall small values for given subjects, which were always pre-focal due to their sentence-initial position, and the consistently large values for given sentence-final objects. Therefore, the model in Table 7 distinguished the two

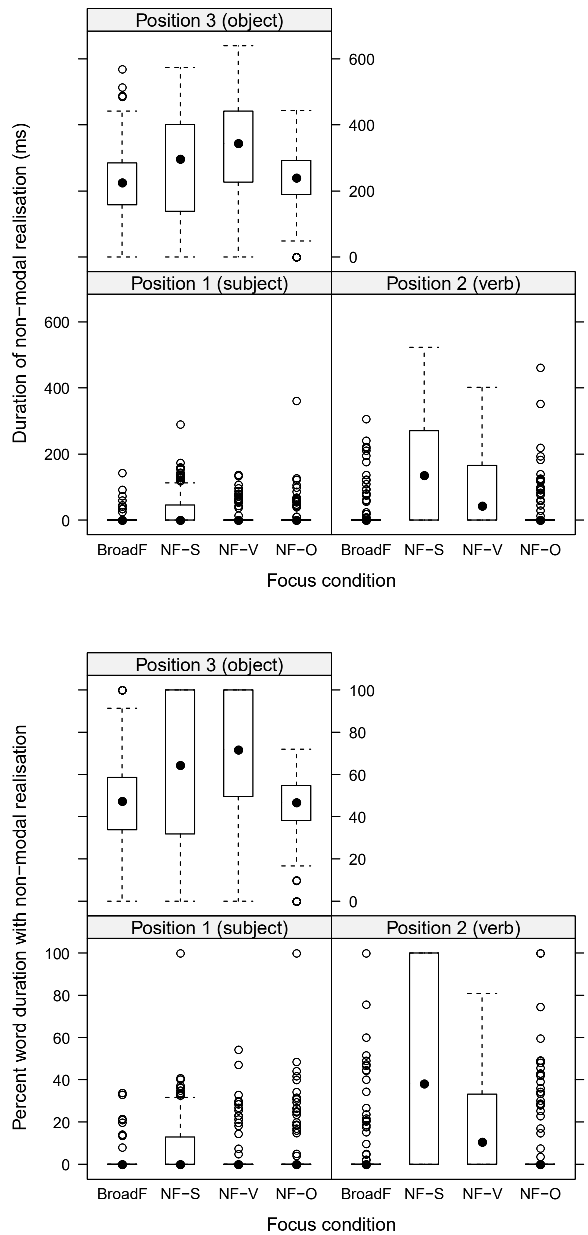


Figure 5: Duration of stretches with non-modal voice quality in ms (top panel) and as percent of word duration (bottom panel) for broad focus (BroadF), narrow focus on the subject (NF-S), narrow focus on the verb (NF-V) and narrow focus on the object (NF-O) in the three positions.

Table 7: Coefficients of the best linear mixed-effects model of duration of stretches with non-modal voice quality (in ms). Random effects: participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	−0.4297	12.0279	−0.0357	0.9716	
Narrow focus (CF)	19.8544	8.8663	2.2393	0.0252	*
Narrow focus (IF)	15.0943	8.8773	1.7003	0.0892	.
Given post-focal (CF)	120.5760	9.1414	13.1901	0.0000	***
Given post-focal (IF)	85.4719	9.1508	9.3404	0.0000	***
Given pre-focal (CF)	−6.1997	9.1414	−0.6782	0.4977	
Given pre-focal (IF)	−1.5199	9.1508	−0.1661	0.8681	
Quantity 2	12.6368	10.5395	1.1990	0.2324	
Position 2	30.5030	8.2065	3.7169	0.0010	**
Position 3	194.7208	8.6707	22.4574	0.0000	***
Narrow focus (CF) : Quantity 2	32.4608	12.5861	2.5791	0.0100	**
Narrow focus (IF) : Quantity 2	37.9762	12.5705	3.0211	0.0025	**
Given post-focal (CF) : Quantity 2	6.1445	12.8993	0.4763	0.6339	
Given post-focal (IF) : Quantity 2	0.9628	12.8738	0.0748	0.9404	
Given pre-focal (CF) : Quantity 2	−9.1914	12.8830	−0.7135	0.4756	
Given pre-focal (IF) : Quantity 2	−8.5479	12.8738	−0.6640	0.5068	

types of givenness and indeed found that the duration of non-modal stretches was significantly larger in post-focal position, but not pre-focally (with the same effects appearing for corrective focus and information focus).[3] The model moreover indicated longer non-modal stretches in narrow focus, although this effect only reached significance for corrective focus and was marginal for information focus. Both focus types showed significant interactions with quantity, suggesting that non-modal stretches were longer in focused words with quantity 2 vowels, which had overall longer durations (cf. section 3.1). Also for both focus types, the estimates associated with the effect of post-focal givenness were larger than those associated with narrow focus. Fig. 5 illustrates this most clearly for the sentence-medial verbs, where durations of non-modal stretches were longer in verb focus than in broad focus, but still mostly covered less than 40% of the word. By contrast, larger values appeared in subject focus, indicating that like for given objects, post-focal verbs were frequently completely realised with non-modal voice quality.

To examine the location of non-modal stretches, each syllable was classified as either completely realised with modal voice or as—partly or completely—non-modal. Fig. 6 depicts the proportion of modal and non-modal realisations by syllable position in the sentence for the different information structures. In each panel, the first stacked bars indicate realisations of the first syllable of the sentence-initial subject, the second ones realisations of the second syllable of the subject, the third ones indicate realisations of the first syllable of the sentence-medial verb, and so on. Generally, non-modal realisations were more frequent towards the end of the sentence, with the final syllable displaying non-modal voicing in over 80% of the cases for all information structural conditions. Crucially, however, this trend was modified by information structure in a consistent

[3] Note that since only verbs occurred in both pre- and post-focally given condition, interactions between the factors focus condition and position could not be tested. However, the fit of the reported model did not differ significantly from more complex models which included this interaction, but did not distinguish pre- and post-focal givenness.

way. For example in narrow verb focus, the bars in Fig. 6 indicate less than 15% non-modal realisations for both syllables of the sentence-initial subject as well as for the first syllable of the focused verb, while all following syllables showed a frequency close to 60% and above. The pattern was similar for subject focus, but with the increase in non-modal realisations appearing on the second syllable of the subject already. The broad focus condition was very similar to narrow focus on the sentence-final object, with both conditions showing low percentages of non-modal realisations up until the first syllable of the object, with a marked increase again appearing on the second syllable of the object. The pattern can be generalised as follows: The average percentage of non-modal realisations increased on the second syllable of a narrowly focus words and all following syllables.

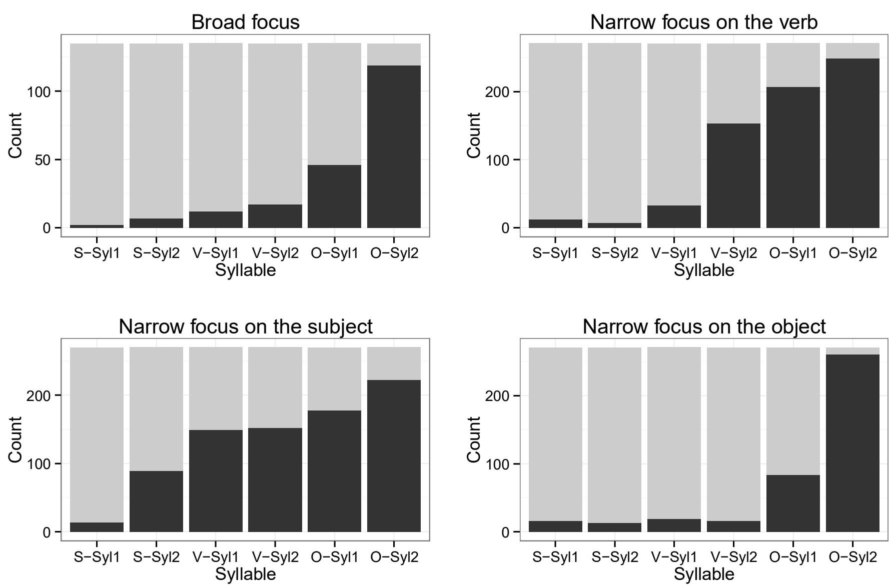


Figure 6: Number of realisations with non-modal voice quality (black bars) and modal voice quality (grey bars) for all syllables in different information structural conditions.

Thus, higher percentages of non-modal realisations generally appeared sentence-finally, on the second syllables of narrowly focused words and on given

Table 8: Coefficients of the best binomial linear mixed-effects model of first syllable voice quality. Random effects: participant, item and trial. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−4.0743	0.4390	−9.2808	0.0000	$***$
Narrow focus (CF)	0.2194	0.2243	0.9782	0.3280	
Narrow focus (IF)	0.0370	0.2282	0.1621	0.8712	
Given post-focal (CF)	3.1316	0.2265	13.8273	0.0000	$***$
Given post-focal (IF)	2.3253	0.2137	10.8837	0.0000	$***$
Given pre-focal (CF)	−0.5145	0.3069	−1.6764	0.0937	.
Given pre-focal (IF)	0.0334	0.2719	0.1229	0.9022	
Position 2 / verbs	0.9380	0.3136	2.9914	0.0028	$**$
Position 3 / objects	2.2678	0.3215	7.0541	0.0000	$***$
Quantity 2	0.9782	0.2364	4.1371	0.0000	$***$

words in post-focal position, while pre-focal words and first syllables of narrow focus words showed few non-modal realisations. To test this interpretation, binomial models of the binary dependent variable—modal vs. non-modal voice quality realisations—were separately fit for first and second syllables. As for the model of the duration of non-modal stretches, models distinguished between pre- and post-focally given words.[4]

The two resulting models confirmed the above generalisation about the effect of focus condition: For first syllables, only post-focal given words showed significantly more non-modal realisations compared to the broad focus condition, while narrow focus and pre-focal given words showed no significant difference (cf. Table 8; positive estimates indicate more frequent occurrences of non-modal

[4]Again, interactions between the factors focus condition and position could not be tested. However, alternative models not distinguishing pre- and post-focal givenness failed to converge when including this interaction.

Table 9: Coefficients of the best binomial linear mixed-effects model of second syllable voice quality. Random effects: participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.6911	0.3026	-8.8938	$0e+00$	$***$
Narrow focus (CF)	1.9434	0.1986	9.7874	$0e+00$	$***$
Narrow focus (IF)	1.7918	0.1969	9.0976	$0e+00$	$***$
Given post-focal (CF)	1.5461	0.2062	7.4995	$0e+00$	$***$
Given post-focal (IF)	1.1974	0.1995	6.0014	$0e+00$	$***$
Given pre-focal (CF)	-1.6388	0.3255	-5.0343	$0e+00$	$***$
Given pre-focal (IF)	-1.0117	0.2693	-3.7567	$2e-04$	$***$
Position 2 / verbs	1.2142	0.2062	5.8889	$0e+00$	$***$
Position 3 / objects	4.0097	0.2437	16.4503	$0e+00$	$***$

realisations in a particular condition). For second syllables, non-modal realisations in given words were also significantly more frequent post-focally (cf. Table 9). Additionally, second syllables of narrow focus words were realised with non-modal voice quality significantly more often, while pre-focal given words showed significantly less non-modal second syllables. Effects were the same for both types of narrow focus, information focus and corrective focus, in both models. Both models also showed effects of position, indicating significantly more non-modal realisations for sentence-medial verbs and for sentence-final objects than for the sentence-initial intercept. Finally, an additional effect of quantity suggested an increase in non-modal realisations for first syllables containing quantity 2 vowels.

To summarise, this section showed that non-modal voice quality appeared more frequently later in the sentence, as well as on words in narrow focus and post-focally given words. For words in narrow focus, non-modal voice quality was generally restricted to second syllables, meaning the durations of non-modal stretches were shorter than for post-focal words, where they frequently covered both syllables.

3.4. Pauses

The two annotators marked altogether 111 pauses in the 947 evaluated sentences based on auditory impression and inspection of waveform and spectrogram. While sentences with noticeable hesitations and slips of the tongue were eliminated altogether (recall section 3), all perceptible pauses were evaluated without applying a minimum duration threshold (see Romøren & Chen, 2015, for a study indicating that this more inclusive method is preferable). The shortest detected pause had a duration of 16 ms, the longest one was 288 ms long, while median duration was 117 ms.

In the short SVO sentences of the present materials, there were two possible positions for pauses to occur: first, between the subject and the verb and second, between the verb and the object. Fig. 7 shows the occurrence of pauses in both positions by information structure, again collapsing the two types of narrow focus for the sake of simplicity. Although numbers were overall small, a clear effect emerged. Between subjects and verbs, i.e. after subjects, pauses were most frequent in narrow subject focus conditions. Pauses were detected after 46% of subjects in corrective focus and after 38% of subjects information focus (33 and 27 cases, respectively). By contrast, only 3 subjects in broad focus and 9 given subjects were followed by pauses (4% and 12% of realisations, respectively). Between verbs and objects, pauses appeared for 46% and 26% of realisations with corrective and information focus on the verb, respectively, and for 28% of realisations with given verbs (18, 10 and 11 cases, respectively). No pauses after verbs were found in broad focus condition.

Thus, pauses were most frequent after words in narrow focus, as confirmed by the model in Table 10 (effects with positive estimates indicate more frequent pause occurrence). The model further suggested that pauses were overall less frequent after verbs than after subjects. There was no significant effect of first syllable quantity on the occurrence of pauses and no indication for an interaction

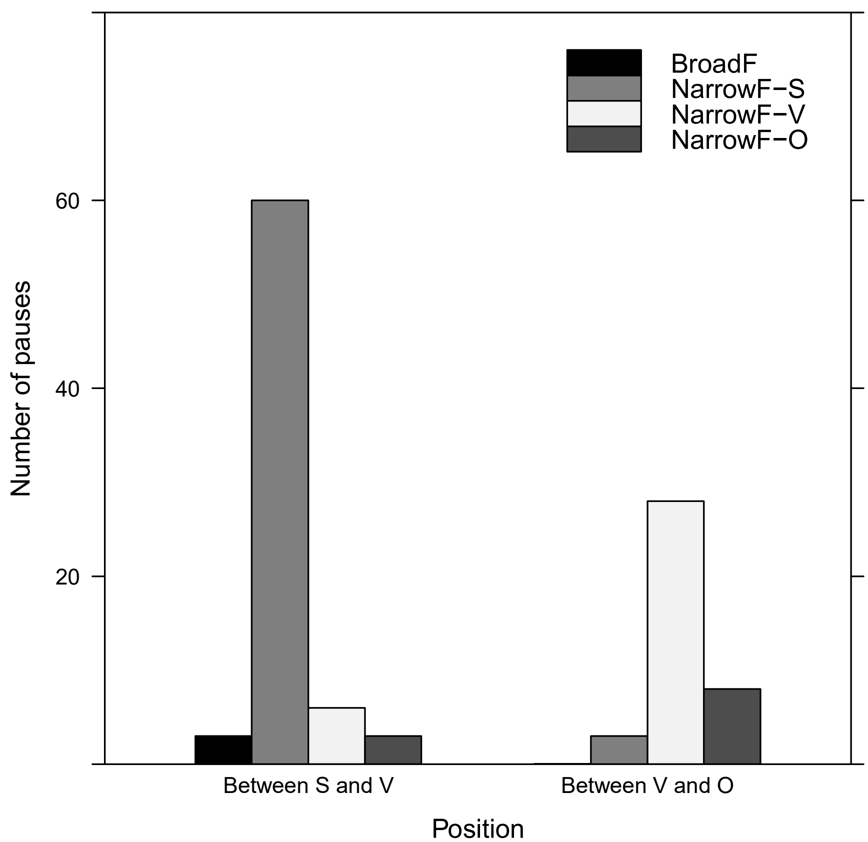


Figure 7: Occurrence of pauses by information structure.

Table 10: Coefficients of the best binomial linear mixed-effects model of occurrence of pauses after subjects and verbs. Random effects: participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 1894$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−6.1410	0.9367	−6.5563	0.0000	$***$
Narrow focus (CF)	3.6878	0.6343	5.8138	0.0000	$***$
Narrow focus (IF)	3.0745	0.6345	4.8454	0.0000	$***$
Given (CF)	0.8089	0.6611	1.2236	0.2211	
Given (IF)	0.1316	0.7091	0.1855	0.8528	
Position 2 / verbs	−0.9786	0.3429	−2.8539	0.0043	$**$
Quantity 2	0.4413	0.3393	1.3004	0.1935	

between the factors focus and condition.[5]

Since the number of pauses was overall low, there were not enough data points for a statistical evaluation of pause duration. Note, however, that the mean durations of pauses after words in broad focus, narrow corrective focus and narrow information focus were 129 ms, 131 ms and 130 ms, respectively. By contrast, pauses not only occurred more rarely after given words, those that were detected had a clearly shorter mean duration of 89 ms.

Thus, this section indicated that pauses were significantly more frequent (and possibly longer in duration) after words in narrow focus than in broad focus, and overall more frequent after subjects than after objects. It found no significant effects of quantity or givenness.

[5] Although quantity did not have a significant effect, it was retained in Table 10, as an alternative model without the factor quantity did not converge. Models including an interaction between focus condition and position did not converge either and did not include significant interaction effects.

3.5. *Intensity*

Intensity minima, maxima and mean intensity were measured for the vocalic nuclei of all syllables in the data set. Since non-modal voice quality is correlated with lower intensity in several languages (Gordon & Ladefoged, 2001), stretches annotated as non-modal were excluded from these measurements. Thus, only the modal part was measured for 718 vowels partially realised with non-modal voice quality, while no intensity measurements could be obtained for 1279 vowels realised with non-modal voice quality throughout (about 13% and 23% of the data, respectively). This section discusses the findings for measurements of mean intensity. Analyses of intensity minima and maxima yielded generally the same results regarding effects of focus condition and position, while showing minor variation in the significance of differences between the quantities.

Fig. 8 depicts mean intensity values of first and second syllable vowel nuclei of subjects, verb and objects in all focus conditions (again collapsing the two narrow focus types for simplicity). Resembling the average f0 values in Fig. 3, it shows a general downtrend over the course of the sentence, which is clearly modulated by information structure. In broad focus, mean intensity smoothly decreased over the course of the sentence, before dropping sharply on the sentence-final syllable. By comparison, words in narrow focus had higher intensity for first syllable vowels, standing out as peaks in Fig. 8. They were followed by a marked intensity drop on the second syllable of the narrow focus word itself and lowered intensity for vowels of all following post-focal syllables.

Like in the analysis of voice quality, statistical modelling was conducted separately for first and second syllable vowels and distinguished between pre- and post-focal given conditions. The results indicated lower intensity later in the sentence, i.e. in positions 2 and 3, as well as significant effects of focus condition for first and second syllables. By contrast, the two quantities did not differ significantly for first syllables, while for second syllables, the effect of narrow focus was significantly amplified for quantity 2. The intensity peaks on first syllables of narrow focus words visible in Fig. 8 were reflected in a significant positive effect of both focus types in the model for first syllables,

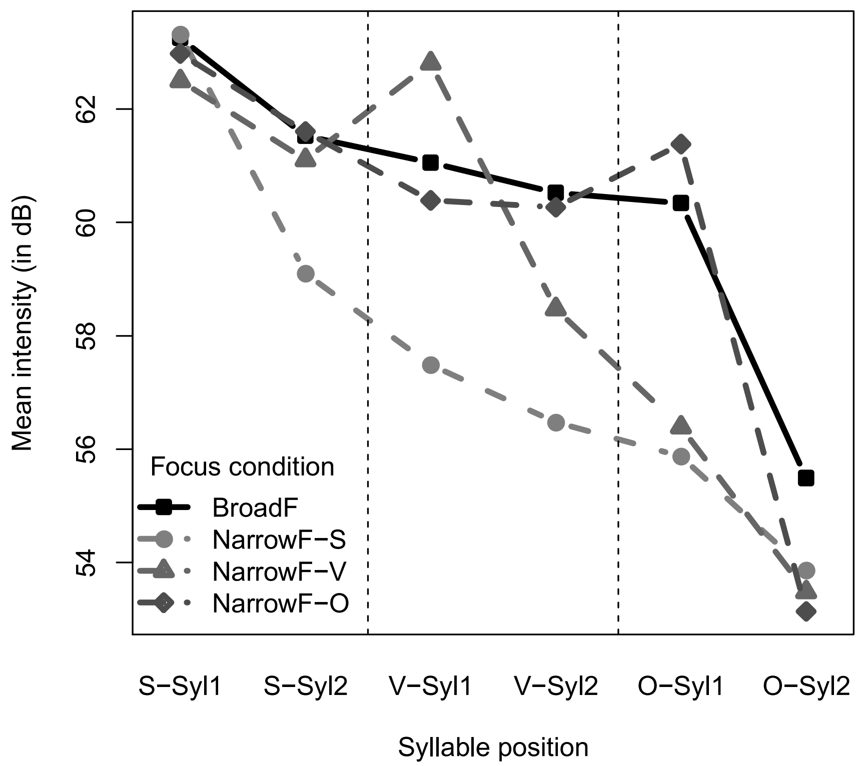


Figure 8: Mean intensity of syllable nuclei by focus condition and position.

Table 11: Coefficients of the best binomial linear mixed-effects model of mean intensity for first syllable vowels (in dB). Random effects: participant and vowel. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2459$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	62.8096	0.8177	76.8111	0.0000	$***$
Narrow focus (CF)	1.1147	0.1534	7.2667	0.0000	$***$
Narrow focus (IF)	0.9210	0.1533	6.0077	0.0000	$***$
Given post-focal (CF)	−4.3162	0.1930	−22.3693	0.0000	$***$
Given post-focal (IF)	−3.1748	0.1816	−17.4800	0.0000	$***$
Given pre-focal (CF)	−0.4609	0.1595	−2.8906	0.0039	$**$
Given pre-focal (IF)	−0.4530	0.1600	−2.8313	0.0047	$**$
Position 2	−1.3107	0.1668	−7.8569	0.0000	$***$
Position 3	−2.7099	0.1736	−15.6069	0.0000	$***$

associated with a larger estimate for corrective focus (see Table 11). By contrast, intensity of first syllable vowels was significantly lower in all given conditions, although this effect was larger post-focally. For second syllables, intensity of pre-focal given words did not differ significantly from the broad focus intercept. However, in narrow focus and post-focal given words, second syllable intensity was significantly lower, in line with the marked drops appearing in Fig. 8.

Thus, this section identified a significant effect of narrow focus, leading to higher intensity on first syllables and lower intensity on second syllables, whereas givenness resulted in lower intensity in post-focal position. It also found effects of position indicating a general intensity downtrend over the course of an utterance, while quantity 2 amplified information structural effects.

4. Discussion

The present study examined the prosodic effects of information structure in Finnish, considering several acoustic parameters. In addition to focus condition,

Table 12: Coefficients of the best binomial linear mixed-effects model of mean intensity for second syllable vowels (in dB). Random effects: participant and vowel. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 1944$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	61.3784	0.7630	80.4424	0.0000	$***$
Narrow focus (CF)	−1.6097	0.2724	−5.9092	0.0000	$***$
Narrow focus (IF)	−1.3642	0.2776	−4.9135	0.0000	$***$
Given post-focal (CF)	−3.0336	0.3642	−8.3287	0.0000	$***$
Given post-focal (IF)	−1.7817	0.3427	−5.1984	0.0000	$***$
Given pre-focal (CF)	−0.1960	0.2611	−0.7506	0.4530	
Given pre-focal (IF)	−0.0767	0.2613	−0.2933	0.7693	
Quantity 2	0.3258	0.3129	1.0413	0.2981	
Position 2	−1.4403	0.1631	−8.8283	0.0000	$***$
Position 3	−4.8924	0.2851	−17.1583	0.0000	$***$
Narrow focus (CF) : Quantity 2	−1.8351	0.4122	−4.4519	0.0000	$***$
Narrow focus (IF) : Quantity 2	−1.1055	0.4182	−2.6433	0.0083	$**$
Given post-focal (CF) : Quantity 2	0.0981	0.5014	0.1956	0.8449	
Given post-focal (IF) : Quantity 2	−0.9798	0.4739	−2.0676	0.0388	$*$
Given pre-focal (CF) : Quantity 2	0.2730	0.3706	0.7365	0.4615	
Given pre-focal (IF) : Quantity 2	0.1776	0.3710	0.4785	0.6323	

it evaluated the effects of position in the utterance (correlated with grammatical role) and (first syllable vowel) quantity. Effects of all three factors appeared, as hypothesised in the introduction. The following sections will discuss these effects in turn and compare them to the hypotheses generated from the literature. It will demonstrate how the present findings relate to previous studies and how they broaden our understanding of information structure marking in Finnish and beyond.

4.1. Effects of information structure

Consistent effects of information structure appeared for most analysed measures: Words in narrow focus had significantly longer durations and higher f0 maxima followed by lower f0 minima, and they were followed by pauses more often than words in broad focus. Conversely, given words had shorter durations, higher f0 minima and (insignificantly) lower f0 maxima. Additionally, the use of non-modal voice quality was significantly more frequent for the second syllables of narrow focus words, as well as for the following post-focal given words. As a result, words in narrow focus and especially post-focal words were realised with significantly longer stretches of non-modal voice quality, while pre-focally given words did not differ from broad focus. Similarly, intensity was significantly higher on first syllables of words in narrow focus and significantly lower on their second syllables and following given words. No reliable difference was observed between the two elicited types of narrow focus, corrective and information focus. While effect sizes differed for several measures, effects of both types of narrow focus consistently pointed into the same direction. This is discussed further in section 4.4; the remainder of this section will simply refer to narrow focus without distinguishing the two types.

In sum, the results showed a wealth of effects complementing and exceeding findings of previous studies of prosodic focus marking in Finnish. They largely confirmed hypotheses regarding duration, f0, intensity and voice quality derived from these existing studies, reproducing several findings within on experiment, while also adding further information.

Confirming hypothesis 1, the present study found longer word durations in narrow focus and shorter ones for given words, in line with Mixdorff et al. (2002); Suomi et al. (2003); Suomi (2007); Nakai et al. (2012). The effect of narrow focus was significant for all segments, matching previous results on Northern Finnish by Suomi et al. (2003) and Suomi (2005, 2007), who observed what they call accentual lengthening to affect all segments from the word onset to the third or forth syllable—although comparing percentages suggests that the increase was consistently smaller in the present study.

In accordance with hypothesis 2, predicting a greater f0 range for words in narrow focus and a smaller range for given words (based on Välimaa-Blum, 1988, 1993; Mixdorff et al., 2002; Suomi et al., 2003; Suomi, 2007; Vainio & Järvikivi, 2007), f0 maxima were higher in narrow focus and lower for given words, while minima were lower in narrow focus and higher for given words. Regarding the competing sub-hypotheses derived in the introduction, the results supported hypothesis 2a. It assumed a high tonal target and a following low tonal target to be affected by f0 range expansions and contractions, based on the analysis in terms of $H_\phi L_\phi$ phrase tones suggested in Arnhold (2014). This is in line with the finding that effects of information structure appeared for L2 measurements (i.e. f0 minima following pitch maxima), but not for L1 measurements (f0 minima preceding maxima). By contrast, this finding contradicts hypothesis 2b, derived from Välimaa-Blum's (1988; 1993) suggestion of L*+H or L+H* accents, which instead predicted effects for preceding minima (L1), but not for following ones (L2). Similarly, the data did not support hypothesis 2c based on Suomi et al.'s (2008) suggestion of LHL accents, which predicted effects for all three measurement points. Lastly, as adjustments affected H and L2 measurements in all positions, f0 range expansions and compressions always altered the size of the pitch fall and not that of the rise. Thus, the study did not support hypothesis 2d derived from studies by Vainio & Järvikivi (2006, 2007), who found that the rise was more important for signalling prominence in sentence-medial position, while the fall was more important sentence-finally.

Regarding intensity, the present study confirmed Vainio & Järvikivi's (2007) finding of a focus effect. Based on their results, hypothesis 3 predicted higher intensity for words in narrow focus and lower intensity for given words. The present data confirmed this hypothesis in measurements of first syllable vowels. By contrast, the sharp intensity drop to the second syllables of narrow focus words visible in Fig. 8 constitutes an effect of information structure that has not previously been described for Finnish. Note that Vainio & Järvikivi (2007) measured intensity at f0 peaks and reported the difference between the values measured at the f0 peaks of two words, finding relatively higher values for words in narrow focus. Since f0 maxima usually appear on first syllable vowels, it makes sense that the current data reproduced their findings with measurements from first syllable vowels. By analysing all vowels for all words in the utterance, the present study provided a broader view of the role of intensity in marking information structure: Narrow focus words did not only stand out via prominent high intensity on their stressed syllables, they also carried a fall that lead to a larger intensity range and was reminiscent of their realisation in final position (see section 4.3).

Moreover, the study confirmed hypothesis 4, assuming a connection between focus and voice quality in Finnish based on a previous study by Vainio et al. (2010). The findings further exceeded the confirmation of the hypothesis, since the predictions for the present data were somewhat unspecific due to the difference in research methodology. Using inverse filtering, Vainio et al. (2010) calculated the glottal air flow for simple SVO sentences in different focus conditions similar to the present ones, finding less tense and more breathy voice quality for narrow focus. Similarly, Airas et al. (2007) found a breathier voice quality for stressed vowels, although with some differences in significance between 'word stress' and 'sentence stress' (their 'sentence stress' condition comprised first-syllable vowels of paragraph-initial content words with high f0; information structure was not manipulated). Their results are difficult to compare to the present ones. In particular, breathy and tense appeared as two extremes in their analysis, corresponding to high and low values of their voice source parameter (normalized amplitude quotient or NAQ), respectively. By contrast, the present analysis employed a binary distinction between modal and non-modal voice. Although different types of non-modal phonation were not distinguished, breathy and creaky voice seemed most common within this category. This impression is in line with the previous literature, which also provides some support for treating breathy, voiceless and creaky voice as belonging to the same category, since all have been observed in the same position and with the same function, marking finality (see section 4.3 for further discussion).

More importantly, however, the two studies differed not only in their analyses, but also in their data. Crucially, Vainio et al. (2010) considered only first syllables in their analysis of the effects of narrow focus. By contrast, the present study evaluated both syllables in disyllabic words and found the most noticeable effect on words in narrow focus to be localised to their second syllables. This effect was quite drastic, with the increased use of non-modal voice quality disrupting f0 tracking to an extent that made measurements impossible (recall the converging sentence-final L2 values shown in section 3.2). Contrarily, Vainio et al. (2010) analysed f0 in parallel to voice quality, discarding only 8% of their data, which showed values below 50 Hz. In fact, they found f0 and NAQ to covary. This suggests that the effect of focus they observed was quite different from the one discovered here. Conceivably, the variation in NAQ values indicating breathier voice for first syllables of words in narrow focus constituted a variation within the range of modal voice. This kind of more subtle variation was undetectable with the present analysis method, but may appear in parallel with the previously-undescribed focus-marking strategy discovered here. Analysing both syllables resulted in further findings not foreshadowed by Vainio et al. (2010): Whereas the use of non-modal voice quality was largely restricted to the second syllables of words in narrow focus, longer stretches of non-modal voice quality appeared for post-focal given words. In these words, both syllables were affected. Furthermore, post-focal given words differed from those in pre-focal position, which did not have longer stretches of non-modal voice quality than words in broad focus condition. Thus, the present data suggest that non-modal voice quality, in addition to its previously-observed functions as a marker of utterance-finality (Lehtonen, 1970; Iivonen, 1998; Myers & Hansen, 2007; Nakai et al., 2009) and as a turn-taking signal (Ogden, 2004), has a further function as a marker of information structure.

Finally, the increased occurrence of pauses after words in narrow focus has, to my knowledge, not previously been reported. No hypothesis could be derived either from studies of Finnish, which have not considered pausing, or from studies of other languages, which show varied results (pre-focal pausing, post-focal pauses or no report of effects on pausing). The present findings thus add to our knowledge of prosodic focus marking in Finnish, identifying another strategy.

Overall, the present study confirmed and extended previous findings. It further distinguished phonological descriptions of f0 movements, favouring an account in terms of $H_{\phi}L_{\phi}$ phrase tones over other existing accounts. Analysing several acoustic measures at once, it found that speakers manipulated all of these measures, employing multiple acoustic correlates in parallel. Although not all effects may appear for every utterance by every speaker, it is compelling that significant effects emerged for five acoustic features—f0, duration, pauses, intensity and voice quality. In fact, information structure influenced all of the phonetic features that are usually characterised as suprasegmental.

4.2. Effects of quantity

As expected, words with quantity 2 vowels had longer durations than words with quantity 1 vowels. Due to morphological restrictions, first syllable vowel quantity varied systematically in all three positions, but second syllable vowel quantity only changed for sentence-medial verbs. Consequently, the effect of quantity on duration was by far largest for verbs, whereas subjects showed the smallest effect. With respect to the difference between subjects and objects, note that the duration of a vowel is not only influenced by the quantity of the vowel itself. Quantity effects on duration are regulated within the first two syllables of a word, so that the structure of a neighbouring syllable has an additional, though smaller effect (see Wiik & Lehiste, 1968; Lehtonen, 1970;

Suomi, 2009; Ylitalo, 2009, on this sub-phonemic variation). This was also observed in the present materials. In particular, analyses of segment durations reproduced studies carried out in the central Finnish variety spoken in Jyväskylä by Lehtonen (1970), who observed that lengthening of the first syllable vowel resulted in a shortening of the second syllable vowel and an inconsistent effect on the intervening consonant if this consonant was phonologically short. The study was not designed to test the interplay of quantity variation of different segments and therefore only analysed effects of the systematically manipulated quantity of first syllable vowels.

The systematic alternation of vowel duration can plausibly account for the other observed effects of quantity. Stretches of non-modal voice quality were longer for words in narrow focus when they contained quantity 2 vowels, and the use of non-modal voice quality on first syllables was overall more frequent in quantity 2. Since quantity 2 vowels were longer, they provided more space for both the realisation of the tonal targets, analysed here as H_{ϕ} and L_{ϕ}, and a non-modal stretch. By contrast, the use of non-modal voice would have risked cutting off the f0 fall for the shorter quantity 1 vowels. Furthermore, higher f0 maxima (H) and lower following f0 minima (L2) appeared for quantity 2 vowels (although the change in L2 did not reach significance as a main effect). Since quantity 2 vowels had longer durations and thus provided more space for f0 excursions to reach the target values, the less extreme values for quantity 1 vowels were likely an instance of compression or undershoot (on accommodation of f0 targets due to lack of phonetic space, see e.g. Grabe, 1998; Fougeron & Jun, 1998; Arvaniti et al., 2006; Hanssen et al., 2007). Similarly, the effect of narrow focus on the intensity of second syllables was also larger for quantity 2 words, where there was more space for the lowering to take effect.

There were no significant effects of quantity for the use of pauses or the voice quality of second syllables.

4.3. Effects of position

This section interprets main effects indicating differences between subjects, verbs and objects in the investigated SVO sentences as effects of position. The introduction predicted downtrends in f0 and intensity over the course of the utterance. In line with this prediction, f0 maxima were significantly lower for both verbs and objects compared to sentence-initial subjects, and the following minima were significantly lower only for the sentence-final objects. This downtrend can be explained as declination, which has been argued to be a universal tendency (Vaissière, 1983) and has previously been observed in Finnish (Iivonen, 1998). Similarly, and also in accordance with the prediction, intensity decreased significantly later in the sentence, a tendency that typically accompanies decreasing f0, although the nature of the relationship between both measures has been debated (Pierrehumbert, 1979; Trouvain et al., 1998; Hird & Kirsner, 2002). Furthermore, words in sentence-final position had significantly longer durations, confirming a hypothesis based on previous studies finding utterance-final lengthening in Finnish (Lehtonen, 1974; Myers & Hansen, 2007; Nakai et al., 2009, 2012). Finally, the data also bore out the last predicted effect of position, as non-modal voice quality was more frequent later in the sentence. This is in keeping with the existing observations of creaky, breathy or voiceless realisations in final positions (Lehtonen, 1970; Iivonen, 1998; Myers & Hansen, 2007; Nakai et al., 2009). Note, however, that a significant increase in the frequency of non-modal realisations appeared already for the sentence-medial verbs in the present materials.

4.4. Focus type

The experimental design elicited prototypical instance of two focus types, corrective and information focus, using yes-no questions with contradictory information and wh-questions, respectively. The difference between prosodic marking of these two focus types has not previously been experimentally analysed for Finnish, so section 1 derived only one tentative hypothesis. This hypothesis was based on findings by Suomi et al. (2003) and their discussion in

Suomi et al. (2008). Suomi et al. (2003) investigated what they termed correlates of stress (no accent), moderate accent and strong accent, and found durational correlates only for their strong accent category. In Suomi et al. (2008), the authors report these results and instead refer to the moderate accents as thematic accents and to the strong accents as contrastive accents, stating that only contrastive accents are marked by durational lengthening, but not thematic ones. Since corrective focus is possibly the most prototypical subtype of contrastive focus (e.g. Krifka, 2008), this generalisation might extend to the present data as well. Thus, hypothesis 2 tentatively suggested that the durational effect of narrow focus may be restricted to contrastive focus and not affect information focus. However, the data did not support this hypothesis, as significantly longer durations appeared for both types of narrow focus. A likely explanation is that words in information focus investigated here did not correspond to Suomi et al.'s (2003; 2008) moderate or thematic accent category (their target words in this category were not narrowly focused and preceded a word marked as emphatic with capitals). In addition to contrastive and thematic accents, Suomi et al. (2008) mention emphatic accents, which they describe as not phonologically distinct from the other categories, and rhematic accents. None of their experimental studies investigated rhematic accents, but their descriptions of the contexts in which they are supposed to occur corresponds to what is termed information focus here. Differences between information focus and contrastive focus could still be predicted if rhematic accents are identified with the realisation of words in narrow information focus, as Suomi et al. (2008) state that unlike the contrastive accent, "the rhematic accent is not normally highly prominent phonetically" (2008, 115), although they do not go into detail.

Different mechanisms of marking contrastive and non-contrastive focus are attested for several other languages. For example in Basque, only corrective focus is marked prosodically (Gussenhoven, 2004), whereas Catalan uses two different pitch accents for narrow contrastive and non-contrastive focus (Prieto, 2014; for a broader discussion of prosodic distinctions between different focus types, see Gussenhoven, 2007). It is also noteworthy that Finnish distinguishes contrastive and non-contrastive focus syntactically, with contrastive elements appearing in initial position and non-contrastive foci appearing sentence-finally per default (Vilkuna, 1989, 1995).

However, the two types of narrow focus, information focus and corrective focus, generally had the same effects in the present materials. While effect sizes frequently differed, significant effects of both types of narrow focus appeared for all acoustic measures. Only for the duration of stretches with non-modal voice quality, the main effect of corrective focus reached significance whereas the same main effect was only marginal for information focus. Otherwise, effects were either significant for both types of narrow focus or not significant for both of them. This suggests that participants applied the same prosodic mechanisms to mark corrective and information focus. However, the difference could lie not in the strategies used, but in the extent to which they are used. Iivonen (1998) distinguishes 'accent for rheme', 'accent for contrast' and 'accent for emphasis', describing them as not necessarily distinct categories, but instead as extremes in a continuum of gradual differences (Suomi et al., 2008, express a similar view for their accent types). In the present data, effect sizes were consistently larger for contrastive focus than for information focus. The analysis did not test whether the difference was significant, but both narrow focus conditions did significantly differ from the broad focus condition, with effects in the same direction. Ultimately, elucidating differences between the prosodic marking of contrastive and non-contrastive focus may require a less controlled and more communication-oriented task.

5. Conclusion

The present study has complemented existing research on the prosodic marking of information structure, extending our knowledge on the basis of a study designed with several methodological considerations in mind.

First and foremost, revealing a complex picture of five acoustic features marking information structure was possible because these different acoustic measures were recorded in parallel. Notably, the results showed a more extensive use of intensity and voice quality than what had previously been observed for Finnish or most other languages. Moreover, studying all measures within one experiment revealed that all of them were affected by information structure and, crucially, that the effects all reached significance. This indicates that speakers adjusted several prosodic parameters at the same time, transmitting information structural differences through several channels in parallel. The study thereby demonstrated that while important cross-linguistic generalisations and differences can be observed when concentrating on f0 and duration, a large part of the picture will be obliterated. The findings of the present experiment thus encourage a broad approach to the study of prosody that takes several acoustic measures into account in parallel.

Second, the study investigated prosodic variation not only for several acoustic measures, but also for all words in the utterances at the same time, again extending the scope of the analysis and providing a detailed picture of prosodic focus marking in Finnish and beyond. Third, this made it possible to take position into account as a factor, finding several interesting interactions for example between the marking of focus and finality. Likewise, the study controlled for lexical quantity, unearthing interactions indicating that effects of information structure were partly more pronounced for words affording more phonetic space for adjustments. Factoring in position and quantity at the same time as varying information structure thus provided a fine-grained picture, showing prosodic variation that was due to information structure, variation that was due to other factors, and the connection and boundaries between them. Finally, a large-enough number of participants and appropriate statistical modelling allowed assessing the significance of effects across the inevitable variation between speakers and items.

In addition to broadening the picture with respect to the investigated acoustic measures and methodology, the findings also made a contribution to a better representation of a wide typological range in prosody research. Thus, the study uncovered several new findings complementing our understanding of focus marking in Finnish, a Uralic language differing from better-studied languages like English genealogically as well as typologically. Thereby, the present study highlights the non-universality of information structure marking, underscoring the fact that a typologically broad data landscape not only means better representation, but also expands and shifts the kinds of phenomena we find.

Acknowledgements

I would like to thank Caroline Féry, Juhani Järvikivi and Martti Vainio for discussions of this data at various stages, as well as Heini Kallio for help in data editing and recording stimulus questions. Mikko Kurimo and Janne Pylkkönen graciously provided the automatic segmentations and Antti Suni wrote a script automatically adjusting them. I am also grateful to Caroline Féry and Anne-Michelle Tessier who provided very helpful comments on an earlier version of this article. The article has also greatly benefited from the criticism of three anonymous reviewers.

This research was supported by Ph.D. grants from the German National Academic Foundation (Studienstiftung des deutschen Volkes) and the Anna Ruths-Foundation.

References

Airas, M., Alku, P., & Vainio, M. (2007). Laryngeal voice quality changes in expression of prominence in continuous speech. In *Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions in Biomedical Applications (MAVEBA 2007)* (pp. 135–138). Florence. URL: http://lib.tkk.fi/Diss/2008/isbn9789512293865/article2.pdf.

Arnhold, A. (2014). *Finnish prosody: Studies in intonation and phrasing*. Phd thesis Goethe University Frankfurt am Main. URL: http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/34798.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*, 667–696. doi:10.1016/j.specom.2005.09.012.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Baayen, R. H., Davidson, D., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. URL: http://linkinghub.elsevier.com/retrieve/pii/S0749596X07001398. doi:10.1016/j.jml.2007.12.005.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (ms). *Parsimonious Mixed Models*. Manuscript submitted to *Journal of Memory and Language*. URL: http://arxiv.org/abs/1506.04967. arXiv:1506.04967.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. URL: http://CRAN.R-project.org/package=lme4 r package version 1.1-10.

Beaver, D., Clark, B. Z., Flemming, E., Jaeger, T. F., & Wolters, M. (2007). When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language*, *83*, 245–276.

Boersma, P., & Weenink, D. (2010). Praat. Doing phonetics by Computer (Version 5.1.29). URL: www.praat.org.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*, 1044–1098. doi:10.1080/01690965.2010.504378.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 27–55). New York: Academic Press.

Epstein, M. A. (2002). *Voice quality and prosody in English*. Phd thesis University of California, Los Angeles.

Féry, C. (1993). *German intonational patterns*. Linguistische Arbeiten 285. Tübingen: Max Niemeyer Verlag.

Féry, C. (2010). Indian Languages as Intonational 'Phrase Languages'. In I. Hasnain, & S. Chaudhury (Eds.), *Problematizing Language Studies. Festschrift for Rama Agnihotri* (pp. 288–312). Delhi: Aakar Books.

Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory*, *31*, 683–734. URL: http://link.springer.com/10.1007/s11049-013-9195-7. doi:10.1007/s11049-013-9195-7.

Féry, C. (2015). *Intonation and prosodic structure*. Oxford: Oxford University Press.

Féry, C., & Ishihara, S. (2009). The phonology of second occurrence focus. *Journal of Linguistics*, *45*, 285–313. doi:10.1017/S0022226709005702.

Fougeron, C. E., & Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics*, *26*, 45–69. doi:10.1006/jpho.1997.0062.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406. doi:10.006/jpho.2001.0147.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. MPI Series 7. Wageningen: Ponsen en Looien.

Grice, M., Baumann, S., & Benzmüller, S. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*, 274–307. URL: http://www.jstor.org/stable/10.2307/416535.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*, 377–417. URL: http://journals.cambridge.org/abstract_S0022226700007799.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Research surveys in linguistics. Cambridge: Cambridge University Press.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 118–145). Oxford: Oxford University Press.

Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus. Cross-linguistic perspectives on meaning and intonation* Studies in Linguistics and Philosophy 82 (pp. 83–100). Dordrecht: Springer. doi:10.1007/978-1-4020-4796-1_5.

Hanssen, J., Peters, J., & Gussenhoven, C. (2007). Phrase-final pitch accommodation effects in Dutch. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1077–1080). Saarbrücken: Pirrot. URL: http://www.icphs2007.de/conference/Papers/1554/1554.pdf.

Hartmann, K., & Zimmermann, M. (2007). In place - out of place? Focus in Hausa. In K. Schwabe, & S. Winkler (Eds.), *On Information Structure, Meaning and Form: Generalizing Across Languages* (pp. 365–403). Amsterdam: Benjamins. URL: http://www.ling.uni-potsdam.de/~mzimmermann/papers/MZ2005InPlace-OutofPlace.pdf.

Hird, K., & Kirsner, K. (2002). The relationship between prosody and breathing in spontaneous discourse. *Brain and language*, *80*, 536–555. doi:10.1006/brln.2001.2613.

Hirsimäki, T., Creutz, M., Siivola, V., Kurimo, M., Virpioja, S., & Pylkkönen, J. (2006). Unlimited vocabulary speech recognition with morph language models applied to Finnish. *Computer Speech & Language*, *20*, 515–541.

Iivonen, A. (1998). Intonation in Finnish. In D. Hirst, & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 311–327). Cambridge: Cambridge University Press.

Iivonen, A., & Harnud, H. (2005). Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association*, *35*, 59–71. doi:10.1017/S002510030500191X.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed model. *Journal of memory and language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007.

Jun, S.-A. (Ed.) (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, S.-A. (Ed.) (2014). *Prosodic typology II. The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*, 243–276.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). *lmerTest: Tests in Linear Mixed Effects Models*. URL: http://CRAN.R-project.org/package=lmerTest r package version 2.0-29.

Ladd, D. R. (2008). *Intonational phonology*. (2nd ed.). Cambridge: Cambridge University Press.

Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish*. Studia Philologica Jyväskyläensia VI. Jyväskylä: Gummerus.

Lehtonen, J. (1974). Sanan pituus ja äännekestot. *Virittäjä*, *78*, 152–160.

Mixdorff, H., Vainio, M., Werner, S., & Järvikivi, J. (2002). The manifestation of linguistic information in prosodic features of Finnish. In *Proceedings of Speech Prosody 2002* (pp. 511–514). Aix-en-Provence. URL: http://www.lpl.univ-aix.fr/projects/aix02/sp2002/pdf/mixdorff-vainio-etal.pdf.

Myers, S., & Hansen, B. B. (2007). The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory*, *25*, 157–193. doi:10.1007/s11049-006-0001-7.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, *37*, 29–45. URL: http://linkinghub.elsevier.com/retrieve/pii/S0095447008000429. doi:10.1016/j.wocn.2008.08.002.

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics*, *40*, 796–807. URL: http://dx.doi.org/10.1016/j.wocn.2012.08.003. doi:10.1016/j.wocn.2012.08.003.

Nevis, J. A. (1986). *Finnish particle clitics and general clitic theory*. Ohio State University Working Papers in Linguistics 33. Columbus, Ohio: The Ohio State University, Department of Linguistics.

Ní Chasaide, A., Yanushevskaya, I., & Gobl, C. (2011). Voice source dynamics in intonation. In W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)* (pp. 1470–1473). Hong Kong.

Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. In E. Couper-Kuhlen, & C. E. Ford (Eds.), *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversations* (pp. 29–62). Amsterdam: John Benjamins.

Ogden, R., & Routarinne, S. (2005). The communicative function of final rises in Finnish intonation. *Phonetica*, *62*, 160–175.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America*, *66*, 363–369. doi:10.1121/1.383670.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Ph.D. thesis MIT.

Prieto, P. (2014). The intonational phonology of Catalan. In S.-A. Jun (Ed.), *Prosodic typology II. The phonology of intonation and phrasing* (pp. 43–80). Oxford: Oxford University Press.

R Core Team (2014). *R: A Language and Environment for Statistical Computing, Version 3.2.2*. R Foundation for Statistical Computing Vienna, Austria. URL: http://www.R-project.org/.

Romøren, A. S. H., & Chen, A. (2015). Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch. *Language and Speech*, *58*, 8–23. URL: http://las.sagepub.com/cgi/doi/10.1177/0023830914563589. doi:10.1177/0023830914563589.

Rooth, M. (1985). *Association with Focus*. Ph.D. thesis University of Massachusetts, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural language semantics*, *1*, 75–116. URL: http://www.springerlink.com/index/10.1007/BF02342617http://www.springerlink.com/index/k57211207j40p176.pdf. doi:10.1007/BF02342617.

Schwarzschild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Natural language semantics*, *7*, 141–177. URL: http://link.springer.com/article/10.1023/A:1008370902407. doi:10.1023/A:1008370902407.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). TOBI: A Standard for Labeling English Prosody. In *2nd Internatioanl Conference on Spoken Language Processing (ICSLP 92)* October (pp. 867–870).

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*, 701–721.

Suomi, K. (2005). Temporal conspiracies for a tonal end: segmental durations and accentual f0 movement in a quantity language. *Journal of Phonetics*, *33*, 291–309.

Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, *35*, 40–55. URL: http://www.sciencedirect.com/science/article/pii/S0095447005000707.

Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics*, *37*, 397–416. doi:10.1016/j.wocn.2009.07.003.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*, 113–138.

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish sound structure. Phonetics, phonology, phonotactics and prosody*. Studia Humaniora Ouluensia 9. Oulu: University of Oulu. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Finnish+Sound+Structure.+Phonetics,+phonology,+phonotactics+and+prosody#0.

Trouvain, J., Barry, W. J., Nielsen, C., & Andersen, O. (1998). Implications of energy declination for speech synthesis. In *SSW3-1998* (pp. 47–52).

Vainio, M., Airas, M., Järvikivi, J., & Alku, P. (2010). Laryngeal voice quality in the expression of focus. In *Proceedings of Interspeech* (pp. 921–924).

Vainio, M., & Järvikivi, J. (2006). Tonal Features, Intensity, and Word Order in the Perception of Prominence. *Journal of Phonetics*, *34*, 319–342.

Vainio, M., & Järvikivi, J. (2007). Focus in production: Tonal shape, intensity and word order. *The Journal of the Acoustical Society of America*, *121*, EL55–EL61. URL: http://link.aip.org/link/JASMAN/v121/i2/pEL55/s1&Agg=doi. doi:10.1121/1.2424264.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler, & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53–66). Heidelberg: Springer.

Välimaa-Blum, R. (1988). *Finnish existential clauses - their syntax, pragmatics and intonation*. Ph.D. thesis The Ohio State University.

Välimaa-Blum, R. (1993). A pitch accent analysis of intonation in Finnish. *Ural-Altaische Jahrbücher*, *12*, 82–94.

Vilkuna, M. (1989). *Free word order in Finnish: Its syntax and discourse functions*. Helsinki: Suomalaisen Kirjallisuuden Seuran Toimituksia.

Vilkuna, M. (1995). Discourse configurationality in Finnish. In K. É. Kiss (Ed.), *Discourse Configurational Languages* (pp. 244–268). New York/Oxford: Oxford University Press.

Wiik, K., & Lehiste, I. (1968). Vowel quantity in Finnish disyllabic words. In P. Ravila (Ed.), *Congressus Secundus Internationalis Fenno-Ugristarum, Helsingiae habitus 23–28. VIII. 1965. Pars I* (pp. 569–574). Helsinki: Societas Fenno-Ugrica/Suomalais-Ugrilainen Seura.

Ylitalo, R. (2009). *The realisation of prominence in three varieties of standard spoken Finnish*. Acta Universitatis Oluensis, B Humanoiara 88. Oulu: University of Oulu. URL: http://en.scientificcommons.org/45956389.

Zerbian, S. (2006). *Expression of information structure in the Bantu language Northern Sotho*. Phd thesis Humboldt University, Berlin. URL: papers://bb7e85db-a986-4eb4-8da5-1804ab10444c/Paper/p546.

Appendix A. Models of segment durations

Table A.13: Coefficients of the best linear mixed-effects model of duration of first syllable onset consonants (in ms) with random effects of participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	68.3152	3.2628	20.9374	0.0000	$***$
Narrow focus (CF)	14.6790	2.5293	5.8036	0.0000	$***$
Narrow focus (IF)	11.7982	2.5261	4.6706	0.0000	$***$
Given (CF)	−4.7286	2.1894	−2.1597	0.0309	$*$
Given (IF)	−9.8837	2.1902	−4.5127	0.0000	$***$
Position 2	4.2801	3.7033	1.1558	0.2535	
Position 3	4.6983	3.7033	1.2687	0.2106	
Quantity 2	−9.0461	3.0268	−2.9887	0.0044	$**$
Narrow focus (CF) : Position 2	6.3352	3.0996	2.0439	0.0411	$*$
Narrow focus (IF) : Position 2	5.4601	3.0962	1.7635	0.0779	.
Given (CF) : Position 2	2.9698	2.6831	1.1069	0.2685	
Given (IF) : Position 2	7.7511	2.6805	2.8917	0.0039	$**$
Narrow focus (CF) : Position 3	−3.2303	3.1023	−1.0413	0.2978	
Narrow focus (IF) : Position 3	−2.9229	3.0933	−0.9449	0.3448	
Given (CF) : Position 3	1.5469	2.6822	0.5767	0.5642	
Given (IF) : Position 3	7.4840	2.6813	2.7911	0.0053	$**$
Narrow focus (CF) : Quantity 2	−3.2787	2.5338	−1.2940	0.1958	
Narrow focus (IF) : Quantity 2	−3.8242	2.5309	−1.5110	0.1309	
Given (CF) : Quantity 2	3.3673	2.1956	1.5336	0.1252	
Given (IF) : Quantity 2	4.6994	2.1942	2.1418	0.0323	$*$

Table A.14: Coefficients of the best linear mixed-effects model of duration of first syllable vowels (in ms) with random effects of participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	78.5486	6.0045	13.0815	0.0000	$***$
Narrow focus (CF)	18.4186	3.9437	4.6704	0.0000	$***$
Narrow focus (IF)	12.9719	3.9437	3.2893	0.0010	$**$
Given (CF)	−6.9621	3.4153	−2.0385	0.0416	$*$
Given (IF)	−9.8983	3.4196	−2.8946	0.0038	$**$
Position 2	−8.3716	6.7999	−1.2311	0.2245	
Position 3	23.4482	6.7999	3.4483	0.0012	$**$
Quantity 2	77.7121	6.8085	11.4140	0.0000	$***$
Narrow focus (CF) : Position 2	6.0791	5.5772	1.0900	0.2758	
Narrow focus (IF) : Position 2	6.6592	5.5877	1.1918	0.2335	
Given (CF) : Position 2	2.3095	4.8300	0.4782	0.6326	
Given (IF) : Position 2	3.4120	4.8330	0.7060	0.4803	
Narrow focus (CF) : Position 3	−11.4217	5.5772	−2.0479	0.0407	$*$
Narrow focus (IF) : Position 3	−9.5881	5.5772	−1.7192	0.0857	.
Given (CF) : Position 3	−3.5393	4.8300	−0.7328	0.4638	
Given (IF) : Position 3	−0.5230	4.8360	−0.1081	0.9139	
Narrow focus (CF) : Quantity 2	27.7568	5.6091	4.9485	0.0000	$***$
Narrow focus (IF) : Quantity 2	24.5319	5.5877	4.3903	0.0000	$***$
Given (CF) : Quantity 2	−18.5525	4.8451	−3.8291	0.0001	$***$
Given (IF) : Quantity 2	−13.9934	4.8451	−2.8881	0.0039	$**$
Position 2 : Quantity 2	6.1313	9.6286	0.6368	0.5274	
Position 3 : Quantity 2	26.1877	9.6286	2.7198	0.0092	$**$
Narrow focus (CF) : Position 2 : Quantity 2	−2.4096	7.9173	−0.3043	0.7609	
Narrow focus (IF) : Position 2 : Quantity 2	7.6692	7.9095	0.9696	0.3323	
Given (CF) : Position 2 : Quantity 2	17.9354	6.8562	2.6159	0.0089	$**$
Given (IF) : Position 2 : Quantity 2	12.0549	6.8498	1.7599	0.0785	.
Narrow focus (CF) : Position 3 : Quantity 2	−10.4553	7.9247	−1.3193	0.1872	
Narrow focus (IF) : Position 3 : Quantity 2	−20.4499	7.9021	−2.5879	0.0097	$**$
Given (CF) : Position 3 : Quantity 2	7.7172	6.8541	1.1259	0.2603	
Given (IF) : Position 3 : Quantity 2	2.5232	6.8519	0.3682	0.7127	

Table A.15: Coefficients of the best linear mixed-effects model of duration of second syllable onset consonants (in ms) with random effects of participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	51.3670	3.7688	13.6295	0.0000	$***$
Narrow focus (CF)	4.9575	0.8382	5.9148	0.0000	$***$
Narrow focus (IF)	4.2082	0.8371	5.0269	0.0000	$***$
Given (CF)	−0.5774	0.7260	−0.7954	0.4265	
Given (IF)	−1.2786	0.7254	−1.7624	0.0781	.
Position 2	15.4355	5.0990	3.0271	0.0058	$**$
Position 3	18.8172	5.0990	3.6903	0.0011	$**$

Table A.16: Coefficients of the best linear mixed-effects model of duration of second syllable vowels (in ms) with random effects of participant and item. CF indicates contrastive focus, IF information focus on the narrow focus constituent. $N = 2841$.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	75.3331	4.9455	15.2326	0.0000	$***$
Narrow focus (CF)	16.4408	2.8957	5.6776	0.0000	$***$
Narrow focus (IF)	12.7041	2.8848	4.4038	0.0000	$***$
Given (CF)	−1.4271	2.5014	−0.5705	0.5684	
Given (IF)	−2.1314	2.5014	−0.8521	0.3942	
Position 2	−11.1986	5.8925	−1.9005	0.0652	.
Position 3	136.2253	5.8925	23.1184	0.0000	$***$
Quantity 2	−18.9914	5.2514	−3.6164	0.0014	$**$
Narrow focus (CF) : Position 2	5.8385	4.0875	1.4284	0.1533	
Narrow focus (IF) : Position 2	4.5167	4.0835	1.1061	0.2688	
Given (CF) : Position 2	2.7534	3.5397	0.7779	0.4367	
Given (IF) : Position 2	0.1743	3.5364	0.0493	0.9607	
Narrow focus (CF) : Position 3	−3.3815	4.0912	−0.8265	0.4086	
Narrow focus (IF) : Position 3	−5.1418	4.0797	−1.2603	0.2077	
Given (CF) : Position 3	−14.7985	3.5386	−4.1820	0.0000	$***$
Given (IF) : Position 3	−11.8002	3.5375	−3.3357	0.0009	$***$
Position 2 : Quantity 2	92.2289	7.4266	12.4187	0.0000	$***$
Position 3 : Quantity 2	−10.8817	7.4266	−1.4652	0.1562	

Appendix B. Alternative models of f0 scaling

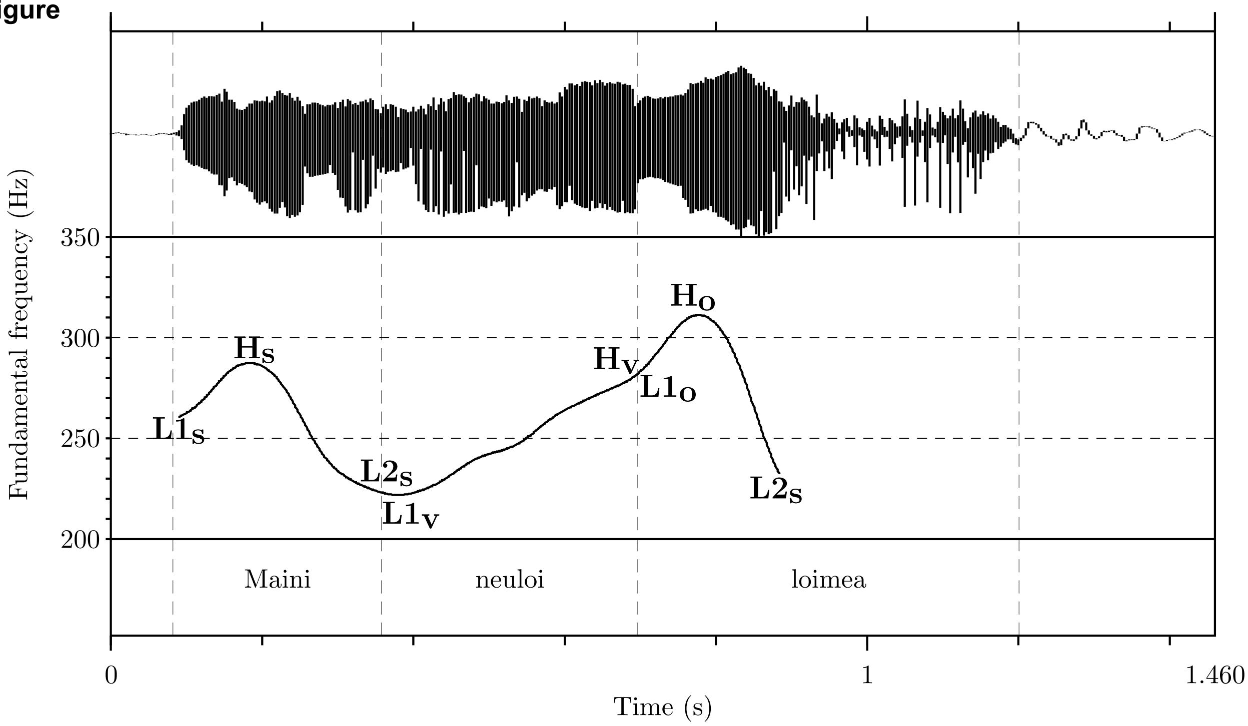
Table B.17: Coefficients of the best linear mixed-effects model of f0 of L1 (in st) not considering neighbouring measurements as predictors. Random effects: participant and item. $N = 2081$.

	Estimate	Std. Error	t value	$\Pr(>\vert t\vert)$	
(Intercept)	25.1308	0.5901	42.5903	0.0000	$***$
Narrow focus (CF)	0.4361	0.3273	1.3325	0.1828	
Narrow focus (IF)	0.6263	0.3332	1.8797	0.0603	.
Given (CF)	−0.2420	0.2970	−0.8148	0.4153	
Given (IF)	0.3904	0.3005	1.2991	0.1941	
Position 2	−2.2142	0.3882	−5.7038	0.0000	$***$
Position 3	−1.8047	0.3803	−4.7449	0.0000	$***$
Quantity 2	−0.3891	0.3789	−1.0270	0.3057	
Male gender	−11.5990	0.8211	−14.1261	0.0000	$***$
Narrow focus (CF) : Position 2	1.2540	0.4758	2.6355	0.0085	$**$
Narrow focus (IF) : Position 2	1.2618	0.4819	2.6182	0.0089	$**$
Given (CF) : Position 2	0.2925	0.4545	0.6435	0.5200	
Given (IF) : Position 2	0.0242	0.4468	0.0543	0.9567	
Narrow focus (CF) : Position 3	0.4663	0.4613	1.0108	0.3122	
Narrow focus (IF) : Position 3	0.6512	0.4671	1.3941	0.1634	
Given (CF) : Position 3	−0.9993	0.4647	−2.1505	0.0316	$*$
Given (IF) : Position 3	−1.2169	0.4530	−2.6863	0.0073	$**$
Narrow focus (CF) : Quantity 2	0.8793	0.4680	1.8790	0.0604	.
Narrow focus (IF) : Quantity 2	1.1933	0.4731	2.5225	0.0117	$*$
Given (CF) : Quantity 2	0.6052	0.4210	1.4376	0.1507	
Given (IF) : Quantity 2	0.1576	0.4217	0.3737	0.7086	
Position 2 : Quantity 2	1.5418	0.5386	2.8624	0.0046	$**$
Position 3 : Quantity 2	0.8820	0.5450	1.6183	0.1071	
Narrow focus (CF) : Position 2 : Quantity 2	−2.3942	0.6632	−3.6101	0.0003	$***$
Narrow focus (IF) : Position 2 : Quantity 2	−3.0768	0.6682	−4.6044	0.0000	$***$
Given (CF) : Position 2 : Quantity 2	−2.4189	0.6223	−3.8868	0.0001	$***$
Given (IF) : Position 2 : Quantity 2	−1.5905	0.6135	−2.5927	0.0096	$**$
Narrow focus (CF) : Position 3 : Quantity 2	−1.5900	0.6638	−2.3954	0.0167	$*$
Narrow focus (IF) : Position 3 : Quantity 2	−1.6079	0.6678	−2.4077	0.0161	$*$
Given (CF) : Position 3 : Quantity 2	−1.2852	0.6463	−1.9886	0.0469	$*$
Given (IF) : Position 3 : Quantity 2	−1.1536	0.6283	−1.8360	0.0665	.

Table B.18: Coefficients of the best linear mixed-effects model of f0 of H (in st) considering neighbouring f0 measurements as predictors. Random effects: participant and item. $N = 1809$.

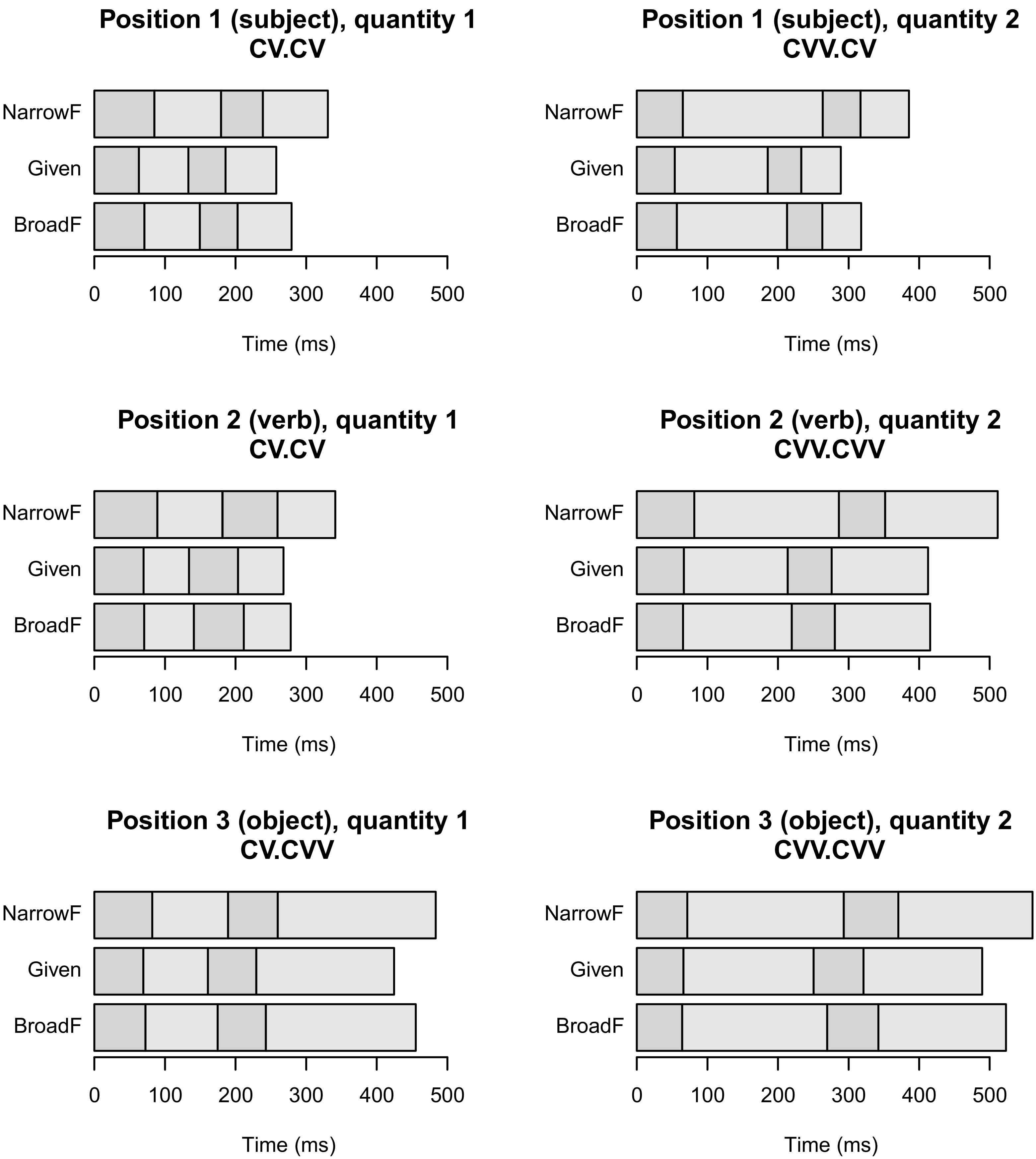
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.9883	0.6191	8.0573	0.0000	$***$
Narrow focus (CF)	1.5489	0.1869	8.2871	0.0000	$***$
Narrow focus (IF)	1.1540	0.1896	6.0857	0.0000	$***$
Given (CF)	−0.6534	0.1712	−3.8167	0.0001	$***$
Given (IF)	−0.5053	0.1703	−2.9663	0.0031	$**$
Position 2	−0.6949	0.2433	−2.8560	0.0050	$**$
Position 3	−0.5463	0.2361	−2.3138	0.0224	$*$
Quantity 2	0.2651	0.1239	2.1406	0.0445	$*$
F0 of L1	0.6071	0.0188	32.3106	0.0000	$***$
F0 of L2	0.2915	0.0181	16.1358	0.0000	$***$
Male gender	−1.0204	0.4643	−2.1974	0.0353	$*$
Narrow focus (CF) : Position 2	0.8197	0.2658	3.0844	0.0021	$**$
Narrow focus (IF) : Position 2	0.7452	0.2688	2.7719	0.0056	$**$
Given (CF) : Position 2	0.2031	0.2603	0.7803	0.4353	
Given (IF) : Position 2	−0.0294	0.2566	−0.1147	0.9087	
Narrow focus (CF) : Position 3	0.2974	0.2579	1.1531	0.2490	
Narrow focus (IF) : Position 3	−0.1389	0.2596	−0.5352	0.5926	
Given (CF) : Position 3	−0.2520	0.2707	−0.9307	0.3521	
Given (IF) : Position 3	−0.5763	0.2545	−2.2644	0.0237	$*$

Figure

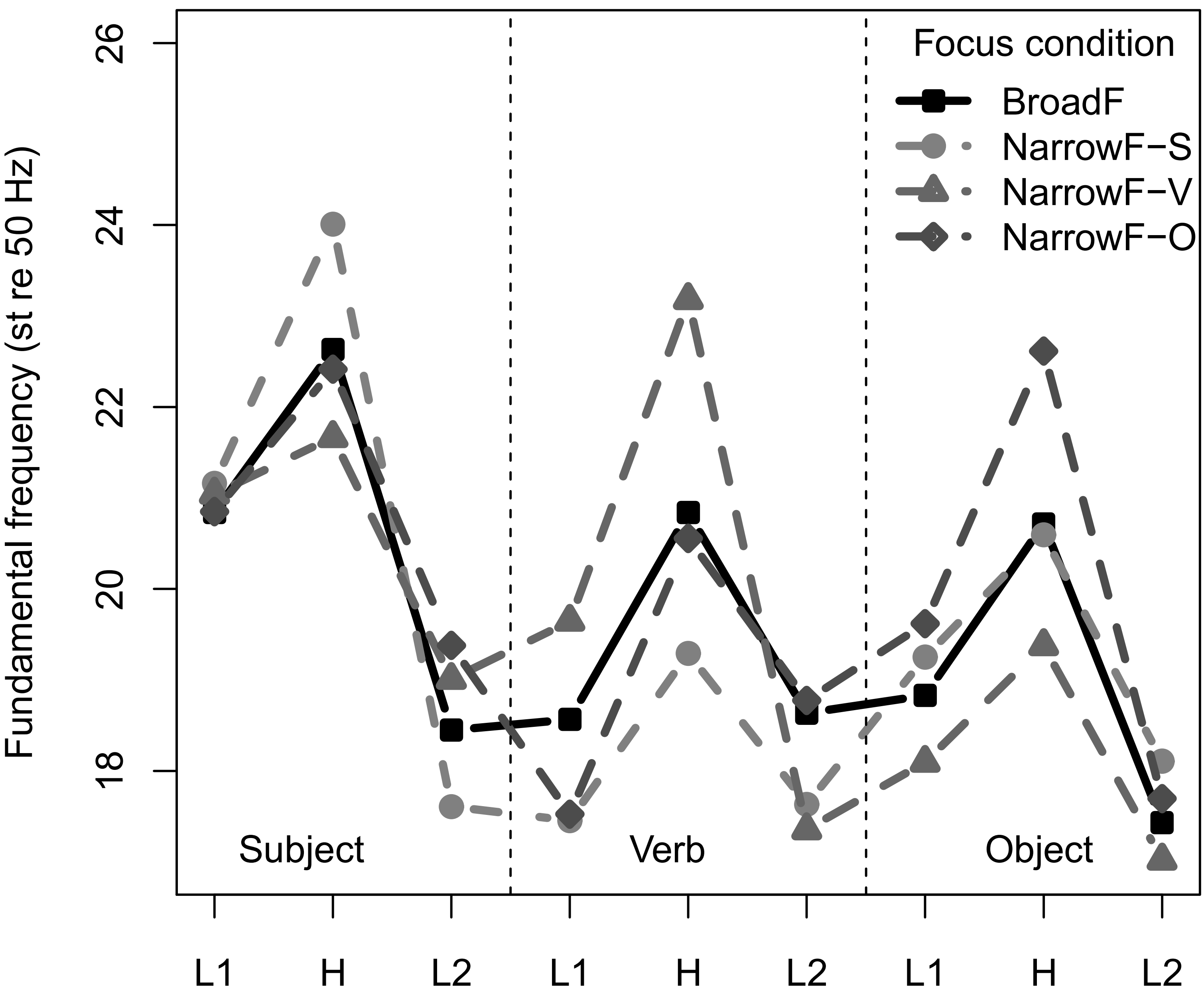


Fundamental frequency (Hz)
350
300
250
200
H_S
$L1_S$
$L2_S$
$L1_V$
H_V
$L1_O$
H_O
$L2_S$
Maini
neuloi
loimea
0
1
1.460
Time (s)

Figure

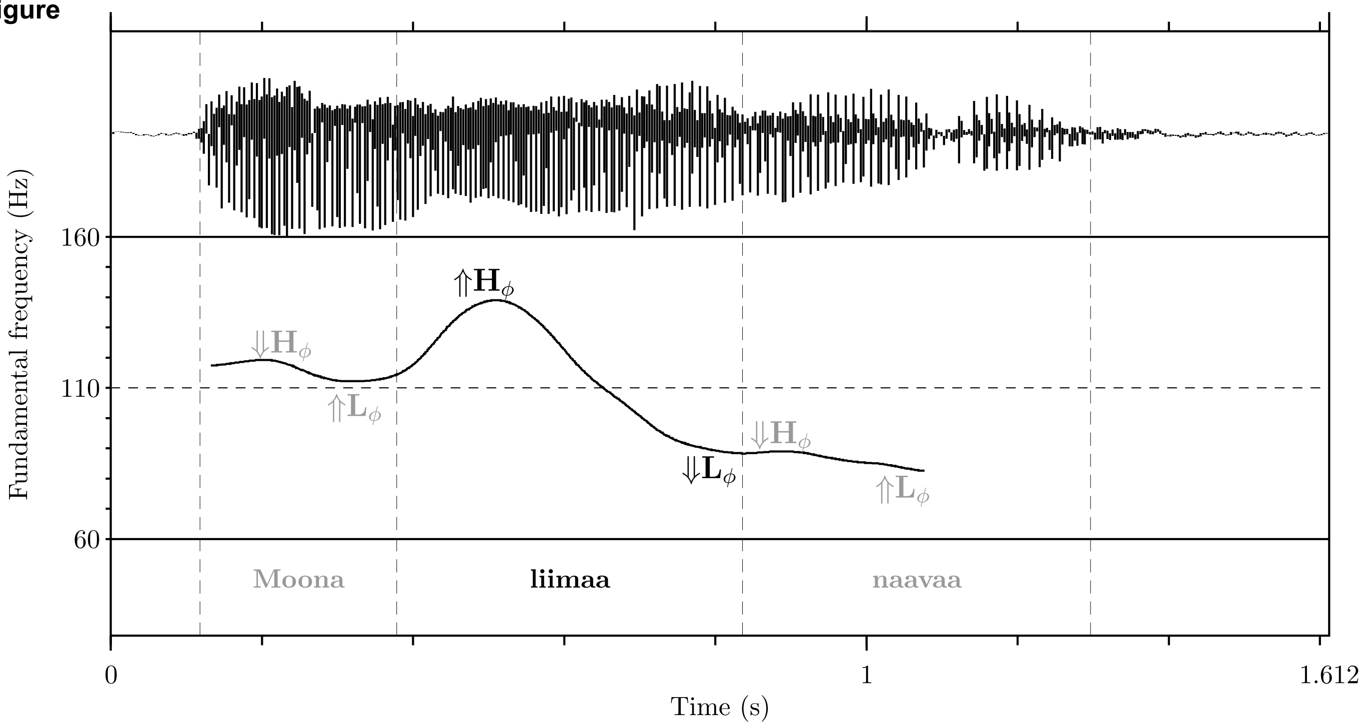

Position 1 (subject), quantity 1
CV.CV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)
Position 1 (subject), quantity 2
CVV.CV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)
Position 2 (verb), quantity 1
CV.CV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)
Position 2 (verb), quantity 2
CVV.CVV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)
Position 3 (object), quantity 1
CV.CVV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)
Position 3 (object), quantity 2
CVV.CVV
NarrowF
Given
BroadF
0
100
200
300
400
500
Time (ms)

Figure

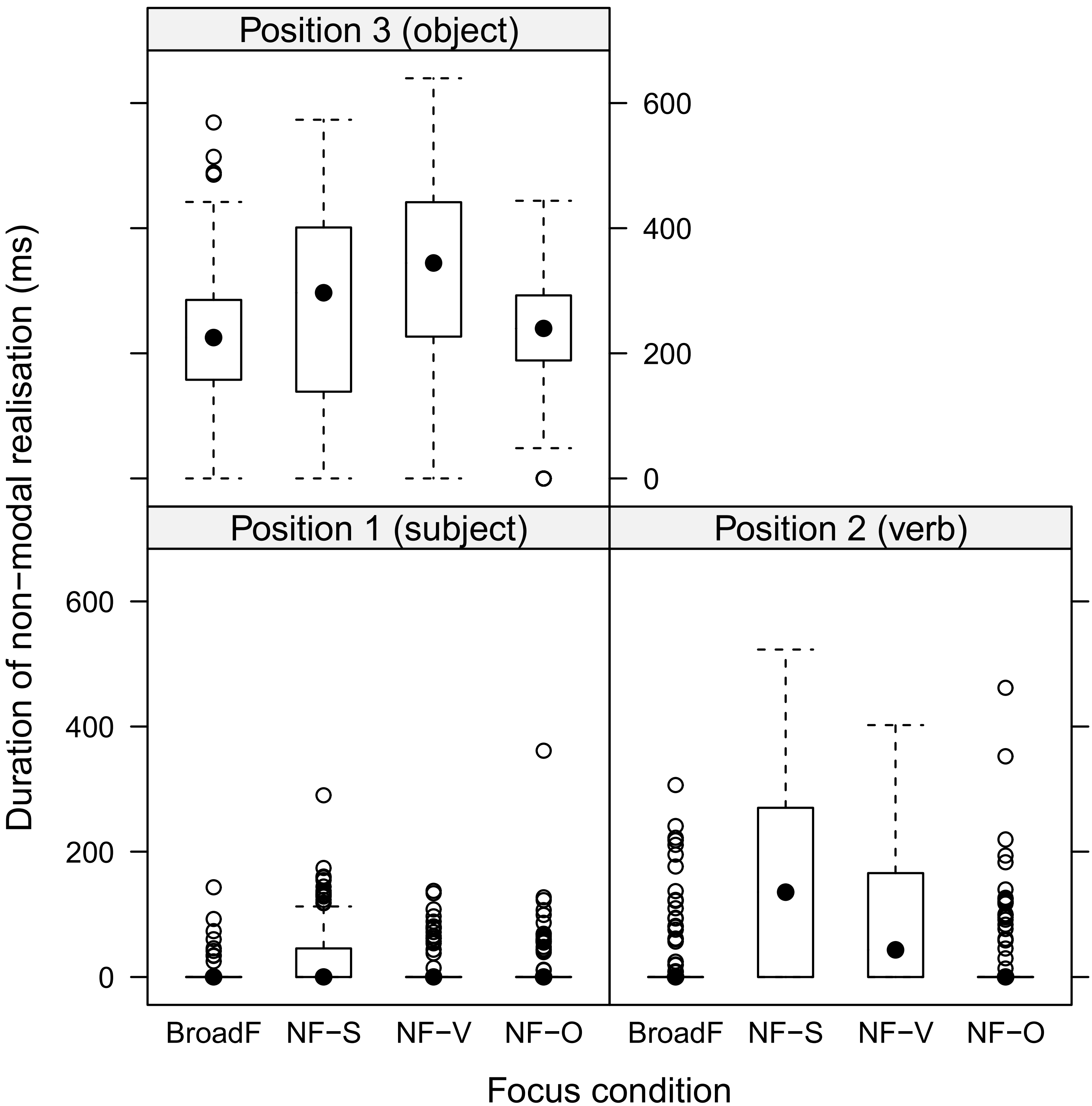


Fundamental frequency (st re 50 Hz)
26
24
22
20
18
Focus condition
BroadF
NarrowF−S
NarrowF−V
NarrowF−O
Subject
Verb
Object
L1
H
L2
L1
H
L2
L1
H
L2

Figure

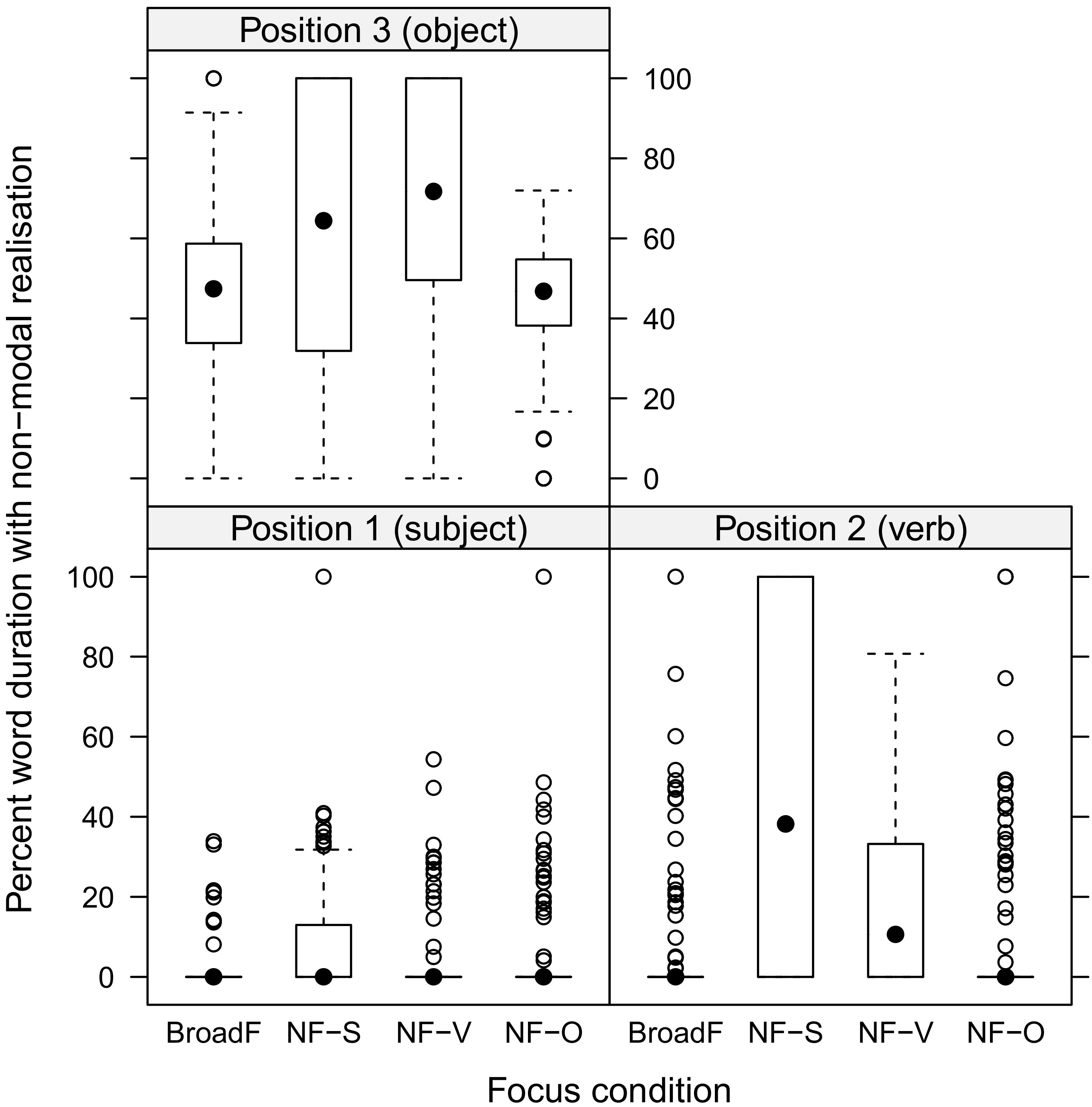

Fundamental frequency (Hz)
160
110
60
⇓H_ϕ
⇑L_ϕ
⇑H_ϕ
⇓L_ϕ
⇓H_ϕ
⇑L_ϕ
Moona
liimaa
naavaa
0
1
1.612
Time (s)

Figure



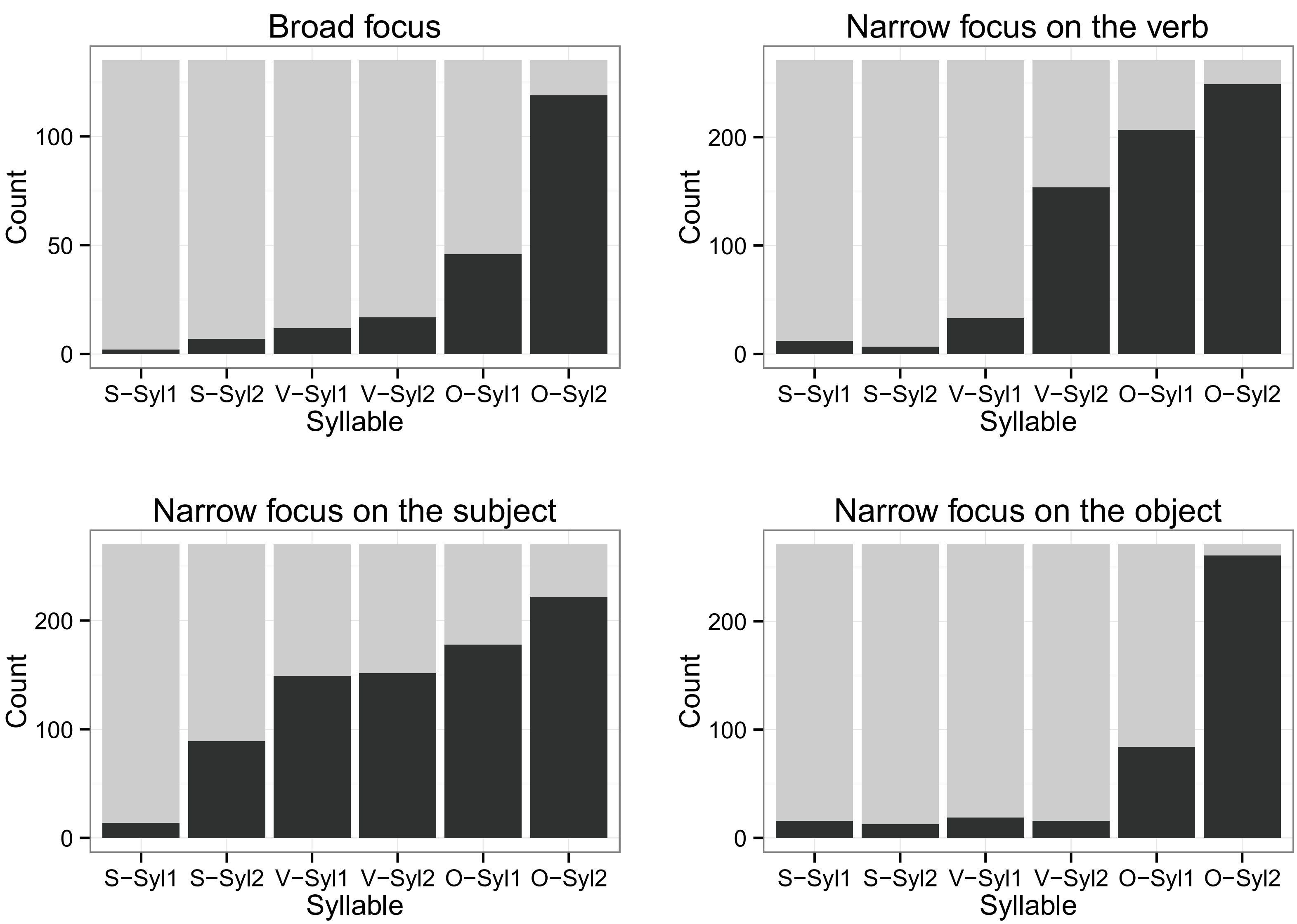
Position 3 (object)
Position 1 (subject)
Position 2 (verb)
600
400
200
0
Duration of non−modal realisation (ms)
BroadF
NF−S
NF−V
NF−O
Focus condition

Figure

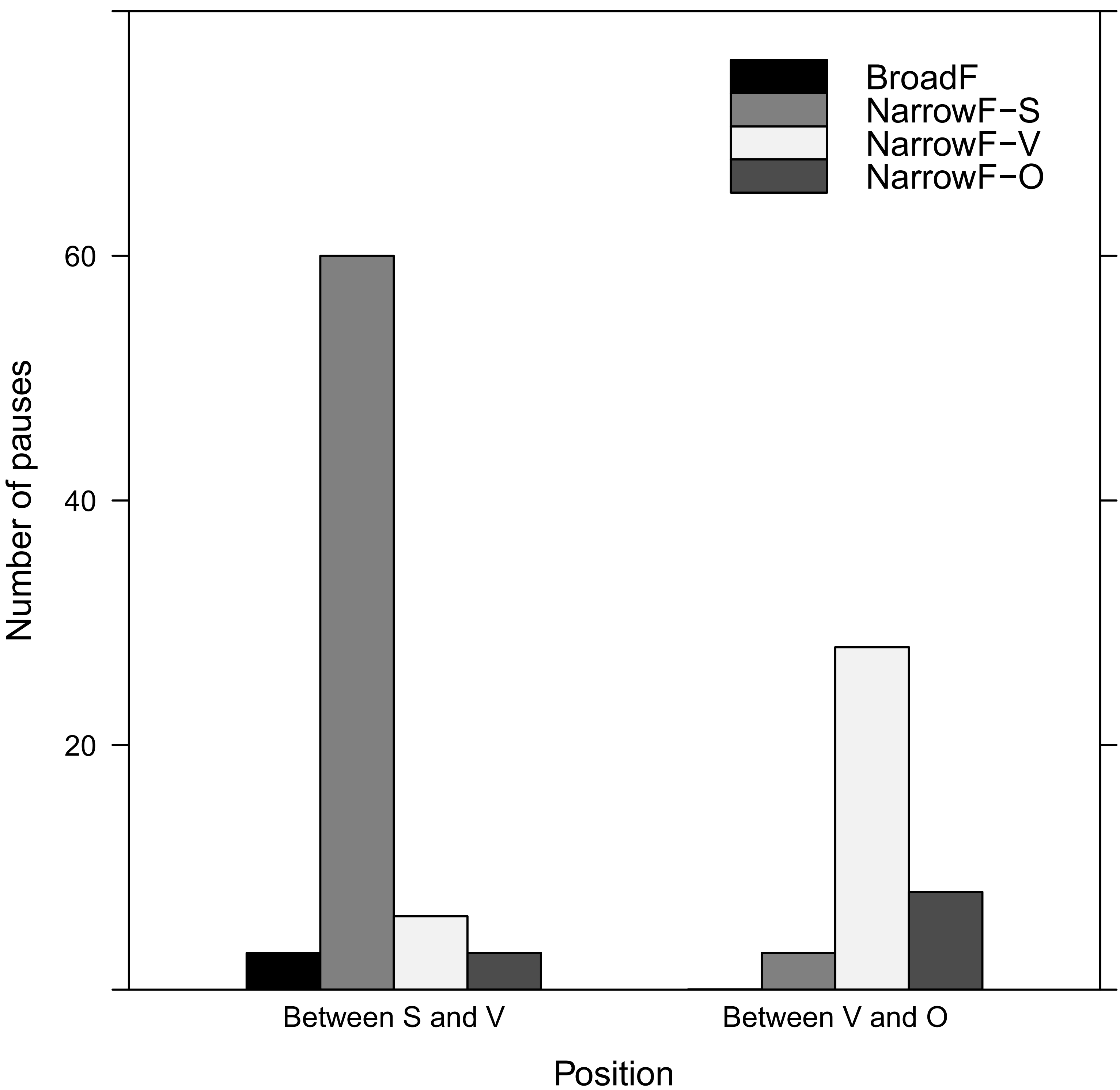


Position 3 (object)
Position 1 (subject)
Position 2 (verb)
Percent word duration with non−modal realisation
0
20
40
60
80
100
BroadF
NF−S
NF−V
NF−O
Focus condition

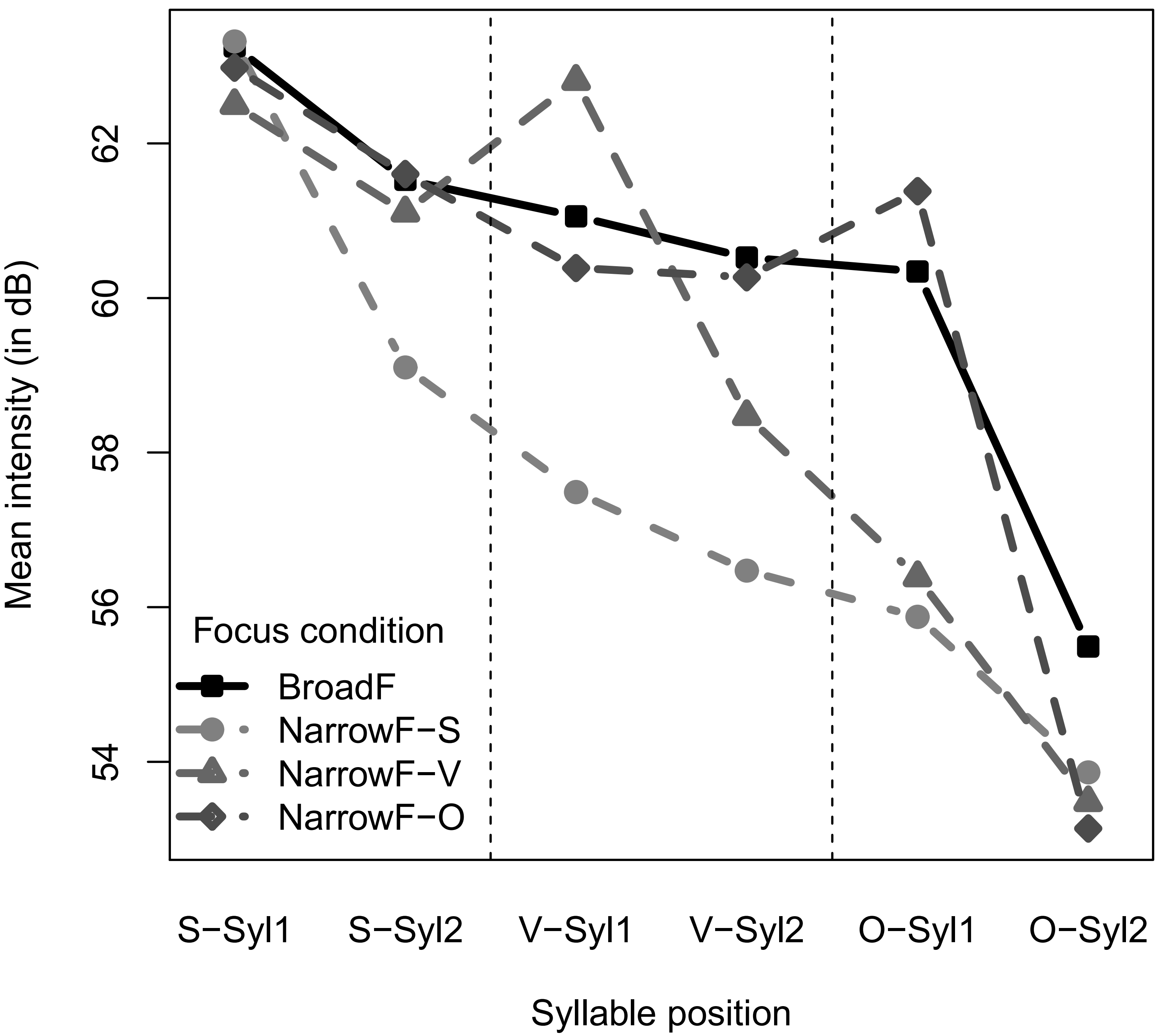
Figure


Broad focus
Count
Syllable
S−Syl1
S−Syl2
V−Syl1
V−Syl2
O−Syl1
O−Syl2
0
50
100
Narrow focus on the verb
0
100
200
Narrow focus on the subject
Narrow focus on the object

Figure


BroadF
NarrowF−S
NarrowF−V
NarrowF−O
60
40
20
Number of pauses
Between S and V
Between V and O
Position

Figure



Mean intensity (in dB)
62
60
58
56
54
Focus condition
BroadF
NarrowF−S
NarrowF−V
NarrowF−O
S−Syl1
S−Syl2
V−Syl1
V−Syl2
O−Syl1
O−Syl2
Syllable position