

# How Truncating Are ‘Truncating Languages’? Evidence from Russian and German

Tamara V. Rathcke

University of Kent, Canterbury, UK

## Abstract

Russian and German have previously been described as ‘truncating’, or cutting off target frequencies of the phrase-final pitch trajectories when the time available for voicing is compromised. However, supporting evidence is rare and limited to only a few pitch categories. This paper reports a production study conducted to document pitch adjustments to linguistic materials, in which the amount of voicing available for the realization of a pitch pattern varies from relatively long to extremely short. Productions of nuclear H+L\*, H\* and L\*+H pitch accents followed by a low boundary tone were investigated in the two languages. The results of the study show that speakers of both ‘truncating languages’ do not utilize truncation exclusively when accommodating to different segmental environments. On the contrary, they employ several strategies – among them is truncation but also compression and temporal re-alignment – to produce the target pitch categories under increasing time pressure. Given that speakers can systematically apply all three adjustment strategies to produce some pitch patterns (H\* L% in German and Russian) while not using truncation in others (H+L\* L% particularly in Russian), we question the effectiveness of the typological classification of these two languages as ‘truncating’. Moreover, the phonetic detail of truncation varies considerably, both across and within the two languages, indicating that truncation cannot be easily modeled as a unified phenomenon. The results further suggest that the phrase-final pitch adjustments are sensitive to the phonological composition of the tonal string and the status of a particular tonal event (associated vs. boundary tone), and do not apply to falling vs. rising pitch contours across the board, as previously put forward for German. Implications for the intonational phonology and prosodic typology are addressed in the discussion.

© 2017 S. Karger AG, Basel

## 1 Introduction

### *1.1 Cross-Linguistic Interplay of Pitch and Segmental Structure*

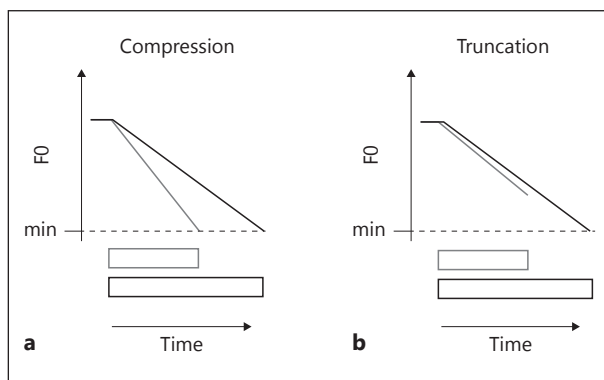
Transmitting pitch information posits two key requirements on segmental structure: it requires (1) time and (2) voicing. Lacking either ingredient is bound to create

difficulties in both production and perception of pitch (e.g., Barnes et al., 2012; House, 2004; Yu, 2010; Zhang, 2002). Particularly compelling cross-linguistic evidence for the interplay of pitch and segments comes from tonal languages where it has been repeatedly observed that the more complex a tone, the longer the voicing of the tone-bearing unit associated with it (Gordon, 2001; 2005; Yu, 2003; Zhang, 2002, 2004). For example, a contour tone cannot be realized on a single mora in Tokyo Japanese (Maeda and Venditti, 1998). Similarly, in Kiowa (Watkins, 1984) and Shan (Morev, 1983), the occurrence of contour tones is restricted to syllables with long vowels and those with a sonorant coda. Based on a survey of 105 languages which have contour tones in their phonological system, Gordon (2001) proposed an implicational hierarchy where long vowels constituted the most likely carriers of complex tones and short vowels the least preferred carriers; the tone-bearing ability of a syllable with a short vowel could be improved by the presence of a sonorant coda consonant. If time and voicing are missing at the level of phonology, phonetically induced lengthening effects have often been observed, like in Cantonese Chinese where contour (in contrast to level) tones induce a lengthening of the tone-bearing vowel (Yu, 2003).

In languages without the phonological tone, the scope of cross-linguistically possible pitch-segment interactions has been far less extensively studied and is thus less well understood. The research does have a long past, but only a relatively short history. Since the 1970s, two strategies have been widely considered to account for all typical modifications of pitch events when time and voicing available for their production are compromised: the intended pitch patterns can either be (a) produced completely in a shorter period of time and therefore compressed, or (b) realized incompletely causing a target undershoot called truncation (figure 1, Eriksson and Alstermark, 1972; Bannert and Bredvad, 1975; Grønnum, 1989; Grabe, 1998). These two strategies, *compression* and *truncation*, were first described in dialects of Swedish, a language which uses pitch accents as part of the lexicon (Eriksson and Alstermark, 1972; Bannert and Bredvad, 1975) and have subsequently been studied in some intonation languages and dialects, adding new evidence that similar pitch modifications can also occur in Dutch, Danish, German, English, Spanish and Catalan (Grønnum, 1989; Grabe, 1998; Grabe et al., 2000; Hanssen, Peters and Gussenhoven, 2007; Prieto and Ortega-Llebaría, 2009). The significance of these phenomena for languages with the postlexical use of pitch has been signified by the proposed typology of truncating vs. compressing languages (Ladd, 1996/2008).

More recent studies, however, suggest that the above dichotomy is quite unlikely to be able to account for all types of time-pressure modifications that happen to the many phonological tone sequences possible in different intonation languages (Grabe et al., 2000; Hanssen, Peters and Gussenhoven, 2007). In particular, these studies highlight the fact that truncation and compression are not mutually exclusive, but can also be found in combination. The use of either strategy may also be controlled by an individual choice rather than a typological setting of a dialect or a language as speakers have an option of lengthening a vowel instead of cutting off the frequency of a pitch target (Prieto and Ortega-Llebaría, 2009).

Truncation and compression have been discussed in the context of phrase-final pitch where time and voicing at the segmental level are controlled by the proximity of the nuclear syllable to the upcoming phrasal boundary, and the amount of pitch information is increased by the proximity of a pitch accent to a boundary tone. At the



**Fig. 1.** Two strategies to cope with the reduced duration of voicing in the accented syllable: **(a)** compression and **(b)** truncation of low pitch targets (both shown in grey) in comparison to non-modified pitch trajectories in black (Eriksson and Alstermark, 1972; Grabe, 1998).

same time, an independent research agenda within the autosegmental-metrical framework used a similar paradigm of variable availability of time and voicing to find an answer to the question of tonal primacy in the phonological form of pitch accents (e.g., Arvaniti et al., 1998, 2000; Caspers and van Heuven, 1993; D’Imperio, 2001; Prieto et al., 1995; Prieto, 2005; Prieto and Torreira, 2007; Schepman et al., 2006; Silverman and Pierrehumbert, 1990; Steele, 1986). Numerous cross-linguistic studies conducted within this framework have shown that tonal crowding or other time-pressure situations commonly cause pitch targets to be adjusted in time, and not only in frequency as the compression/truncation view has suggested.

The starting point of the present study was the idea that the two independent lines of research could be beneficially combined, offering a new perspective on the traditional typology of truncating/compressing languages (Ladd, 1996/2008) and perhaps creating new approaches to the empirical study of phonological forms in intonation languages (Arvaniti et al., 2000).

### 1.2 Truncation and Compression in Russian and German

Pitch patterns in both German and Russian have been identified as showing phrase-final truncation, at least in some phonological environments (Grabe, 1998; Igarashi, 2002; Odé, 2005; Ohl and Pfitzinger, 2009; Rathcke, 2013). Particularly falls seem to be affected by truncation when time and voicing are limited. To address the pitch-segment interactions in German, Grabe (1998) conducted a production study and elicited phrase-final falls (in statements) and rises (in questions), concluding that ‘*German truncates falls but compresses rises*’ (Grabe, 1998:140) in contrast to English, a compression language *par excellence* (cf. Ladd, 1996/2008). In Russian, on the other hand, the rise-falls in yes-no questions are known to be truncated if the sentence ends on an accented syllable (e.g., Igarashi, 2002; Odé, 2005).

Intonation patterns of Russian yes-no questions have attracted research interest for quite some time; this interest is understandable given that there is often no grammatical marker of interrogativity, and pitch acts as the sole carrier of information about sentence mood (e.g., Bryzgunova, 1977, 1980; Igarashi 2002; Makarova, 2003; Odé, 1989, 2005; Svetozarova, 1982; Wenk, 1975). However, little is known about pitch-segment interactions involving other intonational categories of Russian. Similarly, our understanding of the scope of truncation within the intonational

phonology of German is rather limited. Grabe's description of German appears rather minimalistic in comparison to the abundant phonological pitch patterns proposed by GToBI (Grice et al., 2005). In fact, phrase-final pitch falls analyzed as H\*+L 0% in Grabe (1998) correspond to three distinct pitch categories in the GToBI-analysis, namely H\* L-%, H+!H\* L-% and H+L\* L-%. We discussed the empirical basis of the distinction between H+L\* and H+!H\* elsewhere (Rathcke and Harrington, 2010), concluding that it reflects phonetically conditioned pitch variation within H+L\* rather than a genuine phonological contrast. Moreover, GToBI proposes two further nuclear pitch accents, L\*+H and L+H\*, which can be followed by the low boundary tone L-%, thus resulting in a falling phrase-final pitch pattern. Both are rising pitch accents though, and might be subject to compression if Grabe's general conclusion about falls and rises in German applies to accentual, and not boundary-related pitch movements. Given the discrepancy of the phonological analyses and a limited number of falls and rises addressed in previous research (Grabe, 1998), a more detailed investigation is needed to fully understand patterns of phrase-final pitch-segment interactions in German.

Crucially, nuclear rise-falls have been previously studied in both German and Russian using the same categorical perception paradigm (Kohler, 1987, 1991; Rathcke, 2006a, b). In this paradigm, the temporal synchronization of a pitch peak is manipulated through re-synthesis (usually in Praat, Boersma and Weenink, 1997). A stylized production of a nuclear rise-fall is shifted along the time axis in small steps (20–40 ms) resulting in an alignment continuum which is then tested in (1) an identification experiment, to establish how many categories with a distinct meaning can be recognized along the created alignment continuum, and (2) a discrimination experiment, to ascertain that there are indeed clearly defined boundaries between the identified pitch categories. Previous work established that both Russian and German listeners divide such alignment continua into three categories, sometimes termed *early*, *medial* and *late* peaks to reflect the position of the pitch maximum with respect to the accented vowel (Kohler, 1987, 1991; also Grice et al., 2005; Rathcke, 2006a, b among many others). In early peaks, the maximum is located before the accented syllable or vowel which has a falling pitch movement throughout. Medial peaks are characterized by the pitch maximum located within the accented syllable or vowel itself while late peaks show an accentual rise in the vowel with the peak being reached late in the accented syllable or even after. Cross-linguistic meanings of the three types of pitch peaks are summarized in table 1.

Phonological analyses of these three nuclear pitch categories vary in the literature as there are different approaches to autosegmental-metrical representations (cf. Gussenhoven, 2004). Representations chosen as the basis for the cross-linguistic comparisons in table 1 follow the tradition which assumes monotonal and bitonal pitch accents, with the latter being composed of an associated plus a leading or a trailing tone (Ladd, 1996/2008; Pierrehumbert, 1980). GToBI follows these assumptions by and large, and also offers one of the most comprehensive autosegmental-metrical descriptions of German, thus creating a solid basis for this study. The only modification concerns the notation of boundary tones. In the original proposal (Ladd, 1996/2008; Pierrehumbert, 1980), final boundaries of full intonation phrases were assumed to be signaled by a combination of a phrase accent and a boundary tone, as in L-L%. GToBI relinquished the notation of two tonal elements due to the lack of transparency in such representations, reducing L-L% to L-% (and H-L% to

**Table 1.** A summary of nuclear falls in Russian and German

Nuclear pitch trajectory	Phonetic description	Phonological representation		Functional interpretation	
		German	Russian	German	Russian
fall	early peak	H+L* L%	H+L* L%	concluding statement	neutral statement
	medial peak	H* L%	H* (or H*+L) L%	neutral statement	contrastive statement
rise-fall	late peak	L*+H L%	L*+H (or L+H*) L%	surprise, incredulity	yes-no question

H-%, Grice et al., 2005). However, not only the phonetic transparency of the boundary tone notation appears to be problematic, but also the validity of the assumption that an adequate representation of full intonation phrases hinges on the presence of a phrase accent has been questioned, at least in German (Grabe, 1998). Since there exists no empirical evidence that a full account of the Russian or German intonation system requires the presence of a phrase accent, and the exact structure of the prosodic hierarchy in the two languages remains an open question, we decided to reject the annotation of phrase accents and mark boundary tones only (i.e. L%). This seems to be a particularly useful approach in the context of phrase-final truncation where evidence for the presence of even one tone is often difficult to provide, let alone two distinct tonal units.

Unlike German, intonational phonology of Russian has been less well elaborated within the traditional autosegmental-metrical framework. ToRI, a transcription system of Russian intonation (Odé, 2008), is perhaps the most comprehensive account to date, but is based largely on annotations of spontaneous speech, not laboratory work, and follows a deviating perspective on the structure of tonal representations (Gussenhoven, 2004). Nevertheless, some experimental studies do exist and have addressed the three intonational categories relevant to the present study. For example, Makarova (2003) identified L\* L-L% as the typical nuclear accent in declarative sentences, H\* L-L% in exclamations and L+H\* L-L% in interrogatives. The nuclear patterns in declaratives vs. exclamations have also been analyzed as H+L\* vs. H\*+L (Igarashi, 2005) or H+L\* vs. H\* (Igarashi, 2006). Igarashi (2006: 183) further points out the difficulty of correctly identifying the associated tone in the L+H rise, typical of Russian yes-no questions. From an acoustic point of view, both tones seem to be located around the edges of the accented syllable, and the possibility of a contrast between L\*+H and L+H\* is rather doubtful.<sup>1</sup> Igarashi (2006) proposed the L+H\* pitch accent for this type of interrogatives, acknowledging that this proposal is rather arbitrary and therefore unsatisfactory. Complementing evidence from the acoustic studies, the three nuclear pitch patterns were investigated perceptually by Rathcke (2006a, b). A series of perception experiments demonstrated that Russian listeners

<sup>1</sup> Arvaniti et al. (1998, 2000) and Ladd (2004) discuss the issues involved in the decision which one of the two tones within a bitonal pitch accent should be identified as associated when the acoustic signal is ambiguous and does not provide clear cues.

were sensitive to two acoustic cues when categorizing contrastive and interrogative sentences, (1) the slope of the pitch movement and (2) the alignment of pitch peak, but the latter was a much stronger, and thus considered primary, cue to the perceptual contrast leading to the analysis of L\*+H in questions and H\*+L in emphatic statements. Early peaks in neutral statements were analyzed as H+L\*, in agreement with Igarashi (2005, 2006).

As can be seen in table 1, early peaks can be unambiguously analyzed as H+L\* L% and have a similar function in both languages (non-emphatic statements). In contrast, there has been some disagreement with respect to the status of the fall in medial peaks which have sometimes been argued to constitute an accentual fall (i.e. H\*+L) in Russian (Igarashi, 2005; Rathcke, 2006b) and also in German (Grabe, 1998; Uhmman, 1991), with their communicative function differing slightly in the two languages. For the purposes of this investigation, a monotonal pitch accent and a phrasal fall was assumed to be the best representation of medial peaks in the two languages, following the traditional decompositional approach (Ladd, 1996/2008; Pierrehumbert, 1980) and in line with the more recent analyses of this pitch peak type (Grice et al., 2005; Igarashi, 2006). Late peaks are assumed to be best analyzed as L\*+H L% which indicates that the accentual peak is located after the vowel. Their semantic interpretation shows a particular divergence in the two languages: while German late peaks signal surprise or incredulity (Kohler, 1987, 1991; Niebuhr, 2007), Russian L\*+H L% is the typical intonational pattern of yes-no questions (Rathcke, 2006a; Makarova, 2007).

To summarize, the working assumption of the present study is that the early, medial and late peaks in the two languages can be analyzed as H+L\*, H\* and L\*+H, respectively, all followed by the low boundary tone L%. Although previous research has provided sufficient evidence to identify cross-linguistic similarities in the number and the type of pitch categories, semantic differences in the language-specific use of the categories still exist and are unavoidable. However, there is no reason to assume that they will influence phonetic implementation of the pitch categories (Ladd, 1996/2008; Pierrehumbert, 1980), while phonetic implementation of truncation itself can be expected to vary across the two languages (Rathcke, 2013).

### *1.3 Research Questions and Hypotheses*

The main aim of the present study is to provide new evidence on the interplay of pitch and segments in Russian and German, previously classified as ‘truncating languages’ (Grabe, 1998; Igarashi, 2002; Odé, 2005; Uhmman, 1991). The experiments focus on the production of phrase-final falls with an early, medial and late alignment of the high accentual tone, followed by a low boundary tone. In H+L\*, the fall is accentual whereas in L\*+H it is clearly related to the phrasal boundary. For H\*, the most likely phonological interpretation also attributes the falling movement to the phrasal boundary rather than the accentual prominence. The key research questions ask how general a tendency truncation is in the two languages, and if it is sensitive to the status of the low tone, either demarcating a boundary or signaling a pitch accent. If this sensitivity to the phonological status of a tone is indeed the case, a phrase-final nucleus test can potentially be used to inform the debate about the tonal primacy in the phonological form of pitch accents (e.g., Arvaniti et al., 1998, 2000; Caspers and van Heuven, 1993; D’Imperio, 2001; Prieto et al., 1995; Prieto, 2005; Schepman et al., 2006; Silverman and Pierrehumbert, 1990; Steele, 1986).



In addition, the study of German production aims at evaluating the previous conclusion that ‘*German truncates falls and compresses rises*’ (Grabe (1998: 140) and tests if this also applies to the accentual fall in H+L\* and the rise in L\*+H. Against the background of some previous studies (see 1.1), we expect to find evidence for more than just the two strategies to implement tonal information in scenarios of limited time for voicing. Taken together, the results of the production experiments should help to uncover the dynamics of pitch-segment interactions in languages with the post-lexical use of pitch.

Based on the previous research reviewed above, we can expect to observe some pitch modifications to adjust to the deprivation of time and voicing in phrase-final positions. The core working hypothesis of the study assumes that the higher the levels of such deprivation, the stronger the need for pitch adjustments and the more likely they are. The availability of time and voicing can be manipulated by avoiding the presence of a postnuclear syllable (thus making the nuclear syllable the sole carrier of both accentual and boundary-related pitch information) and by changing the segmental composition of the accented syllable from a voiced, sonorant to a voiceless, obstruent environment (thus limiting the possibility of transmitting the relevant pitch information in the nucleus). All of these manipulations would increase the time pressure at the right-hand prosodic boundary, and are expected to affect the immediately adjacent tones, i.e. primarily L%, the preceding H and the relationship between them (which can be expressed as pitch velocity, according to the truncation/compression model). It is also possible that a substantially increased time pressure (e.g. in a phrase-final nucleus flanked by voiceless obstruents) may even extend its scope to L\* (in L\*+H). Overall, late peaks (L\*+H L%) are expected to be particularly prone to large re-adjustments, given that the amount of tonal information is relatively high, and located late within the nuclear accented syllable, closer to the phrasal edge. In contrast, medial (H\* L%) and early (H+L\* L%) peaks may be less affected by the time and voicing deprivation from the right-hand phrase boundary, given a lower pitch information density and the peak location further away from the time pressure source.

Table 2 outlines the relationships between the three pitch parameters of interest, the expected pitch adjustments under an increasing time pressure condition and the corresponding adjustment strategy predicted by the compression, truncation and re-alignment model. The use of these strategies is expected to be specific to each language and pitch accent type, though precise hypotheses are impossible to formulate at this point, due to lack of evidence. However, we can expect a combination of compression, truncation and re-alignment to be observed more frequently than the sole use of one selected strategy.

## 2 Method

### 2.1 Data Elicitation

The study utilized the methodological approach of a pitch imitation technique developed by Pierrehumbert and Steele (1987, 1989). In their study, speakers of American English were presented with a series of f0-peak shift continua and were given the task of imitating the pitch of the phrase ‘*Only a millionaire*’ as closely to the perceived pattern as possible. The authors’ assumption that the imitations would not contain the gradual variability of peak delays but cluster around two categorical pitch accents from the created continua was borne out by the collected production data, and their

**Table 2.** Outline of the predictions of the pitch adjustment model to be tested in the cross-linguistic study

Parameter	Adjustment	Strategy
1. velocity of pitch change	higher velocities under increased time pressure; slower velocities in decreasing time pressure	compression
	stable velocities or variability independent of an increased/decreased time pressure	no compression
2. pitch target scaling	undershoot of a low or a high pitch target under increased time pressure; full target realization in low time pressure	truncation
	stable scaling of pitch targets or scaling variability independent of an increased/decreased time pressure	no truncation
3. pitch target alignment	systematic, continuous temporal leftward shift of the pitch targets with increasing time pressure from the right-hand prosodic boundary	re-alignment
	pitch targets stably anchored in time with respect to the segmental structure or alignment variability independent of an increased/decreased time pressure	no re-alignment

method successfully employed for experimental research in intonational phonology ever since (cf. Gussenhoven, 1999).

The original implementation of the method was modified for the purpose of this study. Instead of listening to synthetical continua, the study participants were exposed to stylized, prototypical realizations of the three pitch accents under investigation (H+L\*, H\* and L\*+H followed by L%). The task in this case involved a perceptual identification of the corresponding pitch pattern and its subsequent reproduction in target words containing different segments and variable syllable structures. The idea behind this imitation experiment is based on the assumption that, faced with the variability of target words and an unambiguous clarity of the pitch category, speakers will need to employ a language-specific pitch adjustment strategy to express the same linguistic meaning (see 1.2) in different linguistic materials (see 2.2.1). Please note that the imitation technique developed for the purposes of this study is remarkably different from the shadowing tasks commonly used in social convergence paradigms where participants are either asked to repeat the exact same words after a model speaker (e.g., Goldinger, 1998; Babel, 2012) or are explicitly instructed to imitate the pronunciation of the exact same words after a model speaker (Adank et al., 2010; Dufour and Nguyen, 2013). However, the method of this study and the abovementioned paradigms are guided by the core assumption that *‘imitation is an all-pervading process by which individuals adjust to one another in social interaction, and is seen as one of the fundamental mechanisms of human development’* (Dufour and Nguyen, 2013:1; cf. Meltzoff et al., 2009).

Pilots confirmed that naive participants benefitted from the task that did not require any explicit meta-linguistic judgment, and the amount of data lost to unintended pitch accent productions was minimal in comparison to the usual data elicitation task involving reading of prepared texts where target words and sentences were embedded in different contexts biasing toward a particular interpretation. Moreover, the data elicitation method allowed to control for such idiosyncratically variable factors like degree of emphasis (e.g., Ladd and Morton, 1997; Liberman and Pierrehumbert, 1984) and personal involvement (e.g., Crespo-Sendra et al., 2010; Granström and House, 2005), thus increasing the comparability of the resulting pitch properties across the dataset.



**Table 3.** Overview of the target words chosen for Russian and German

Stress placement	Consonant type			
	Russian		German	
	Sonorants	Obstruents	Sonorants	Obstruents
penult	<i>Калин</i> кин /кА'linkin/	<i>Калин</i> кин /кА'jɪfkin/	<i>Linner</i> /'lɪne/	<i>Schiffer</i> /'ʃife/
ultima	<i>Жаклин</i> /ʒAk'lin/	<i>Раши</i> в /тА'ʃɪf/	<i>Linn</i> /lɪn/	<i>Schiff</i> /ʃɪf/

2.2 Materials

2.2.1 Experimental Sentences

Experimental materials for this study followed the design of previous cross-linguistic experiments conducted by Grabe (1998). Target words were personal names, embedded in syntactically comparable carrier sentences of the type ‘*It was (Mr) X*’ (‘*Das war Herr X*’ in German and ‘*Eto byl(a) X*’ in Russian) with the target word being nuclear accented and occurring in the phrase-final position. The structure of the target word (*X*, a personal name) was varied with respect to the two characteristics: (1) the type of consonants abutting the stressed vowel within a CVC syllable (obstruents vs. sonorants) and (2) the placement of the stressed syllable within the target word (in the ultima vs. in the penult). This design enabled a systematic manipulation of the duration of voicing available for the realization of the nuclear pitch pattern. Overall, the materials comprised of two trisyllabic, four disyllabic and two monosyllabic words, all of which occurred naturally in Russian and German, and yet had a cross-linguistically comparable segmental structure of the relevant nuclear/postnuclear region. While vowel length is not contrastive in the phonology of Russian (e.g., Bondarko, 1998), tense /i/ and lax /ɪ/ are phonemic in German (e.g., Kohler, 1995). Given that tense vowels tend to receive a considerable amount of lengthening under accentuation (Mooshammer and Geng, 2008), German target words contained only lax vowels to allow for some cross-linguistic comparability of segmental durations. Experimental target words of this study are given in table 3. Please note that all of the Russian test words had a pretonic syllable while the German test words started with a stressed syllable but were preceded by an unstressed monosyllabic word ‘*Herr*’ (Mr), resulting in a similar metrical structure in the materials of the two languages. The syntactic structure of the carrier sentences was identical.

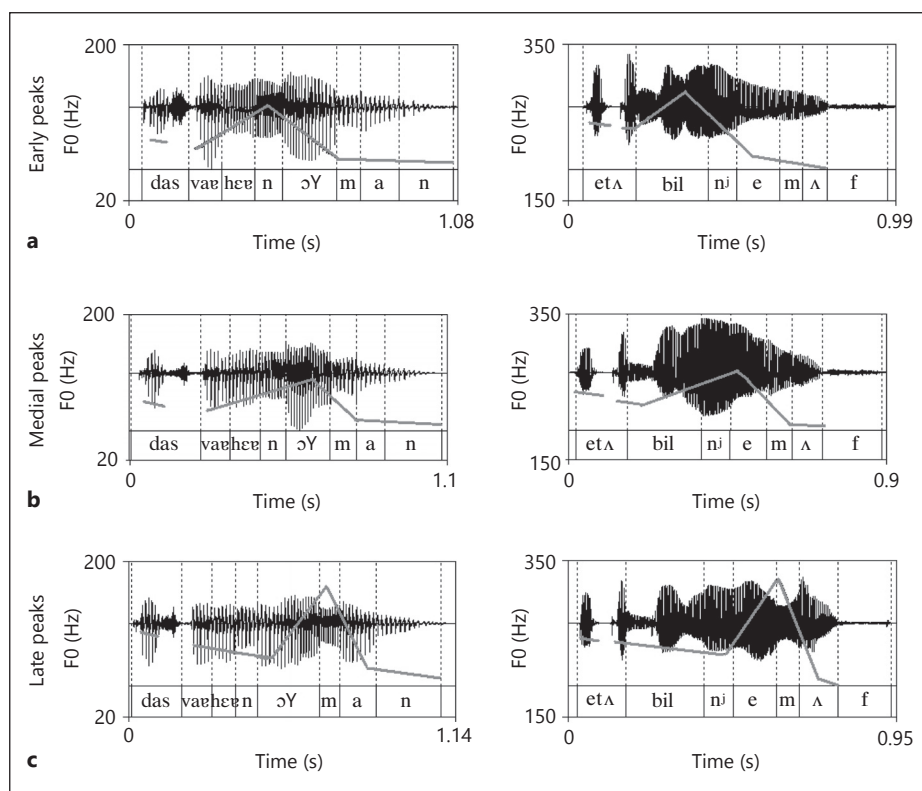
2.2.2 Auditory Prompts

Following sentences served as bases for auditory prompts: (1) ‘*Это Был Немов*’ /etA bil 'niemAf/ in Russian (English: *It was Nemov*) and (2) ‘*Das war Herr Neumann*’ /das vAr hɛr 'noʏman/ in German (English: *It was Mister Newman*). In line with the design of experimental items, the nuclear accented word was a personal name: *Nemov* and *Neumann* are common personal names in the respective language, and also share some phonemic similarity. Both words contain voiced material within and after the stressed syllable and are therefore most likely to carry prototypical, unmodified realizations of the intended pitch patterns. An expert native speaker (36-year-old male German and 30-year-old female Russian<sup>2</sup>) produced the respective base sentence with three intonational structures, i.e. early peaks H+L\* L%, medial peaks H\* L% and late peaks L\*+H L%. These productions are shown in figure 2.

2.3 Procedure and Participants

The target words were presented in Russian or German orthography, respectively, printed on small paper cards. The target words were supplemented by an equal number of fillers. Each word was repeated three times, and the cards were randomized (successive repetitions of the same word were avoided). The auditory prompts were presented through high-quality headphones. Participants were

<sup>2</sup> The author of the paper.



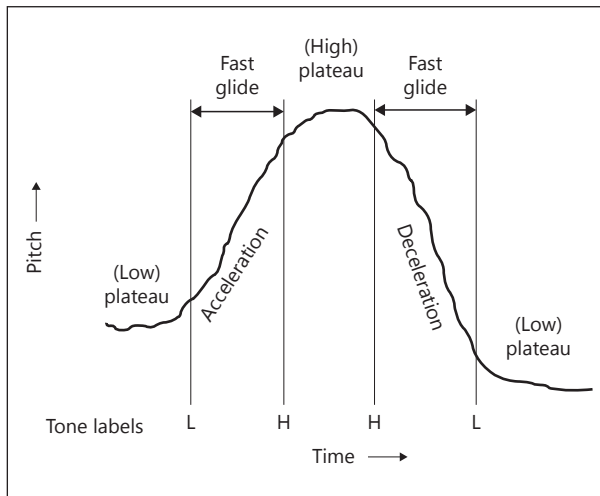
**Fig. 2.** Sound waves with superimposed stylized pitch contours of the German (left-hand panels) and Russian (right-hand panels) auditory prompts, produced as early (panel **a**), medial (panel **b**) and late (panel **c**) peaks.

instructed to listen to the melody of the prompt and to reproduce the same melody in the sentence containing a different personal name, i.e. the one printed on each card.

The recording was split into three sessions. Each session was dedicated to one of the three intonational structures under investigation (i.e. medial, early and late peaks), and started with a brief explanation of the meaning of the pitch pattern to be imitated, with contextualized language-specific examples. In the subsequent training phase, participants repeated the auditory prompt five times, and after they mastered the intended pitch, the experimental recording began. The production of each sentence was preceded by the corresponding auditory prompt, and re-recorded only if a speaker expressed the wish to do so or in case of a slip of the tongue. The author closely monitored imitations during the recordings. In general, speakers of both languages did not experience any difficulties with the procedure. The data of those who showed problems with the task were excluded from the analyses (three German and two Russian speakers).

Productions of 10 Russian (3 male, mean age 31 years) and 10 German (3 male, mean age 25 years) speakers were analyzed. None of the speakers had any speech or hearing disorders. The German participants were from the north of Germany (Schleswig-Holstein area and Hamburg) and lived in Kiel at the time of the recording. The Russian participants were born either in Moskow or St. Petersburg and spoke in standard Russian accent. They were either living in or visiting Munich when the recording took place.

The German experiments were conducted in a sound-insulated booth at the Institute of Phonetics and Digital Speech Processing in Kiel. The recordings of the Russian speakers were made in the soundproofed booth of the Institute of Phonetics and Speech Processing in Munich.



**Fig. 3.** A stylized representation of several stages of pitch production (low/high plateaus and fast glides of acceleration and deceleration) showing L and H turning points as annotated in the present study.

#### 2.4 Data Preparation

The data were manually segmented and annotated using *EMU Speech Database* (Harrington, 2010). Segmentation and labeling decisions were based on the acoustic information (waveform, spectrograms and pitch trajectories) and the auditory impression. The onset and offset of accented syllables were labeled along with the start (H) and the end (L) of phrase-final pitch falls, plus the start (L) and the end (H) of pitch rises for L\*+H pitch accents to capture the accentual pitch movement.

Our approach to the annotation of pitch for this study is illustrated in figure 3. As established in previous research into physiological limits of pitch production (Ohala and Ewan, 1973; Sundberg, 1979; Xu and Sun, 2002), a complex pitch pattern can be decomposed in several stages which, for a fall, includes a deceleration phase, a fast glide and ends in a low plateau (cf. fig. 3). The glide is known to constitute approximately 75% of the overall pitch change and excludes the absolute maxima and minima involved in the production of a pitch pattern. However, the glide gives a more adequate representation of the actual speed of pitch change (Xu and Sun, 2002), and was considered an essential window into the patterns of compression. If the phrase-final low target was completely cut off in frequency, H and L labels were placed next to each other, close to the end of the corresponding pitch track. In case of intervening voiceless consonants (like in target words *Schiffer*, *Кашишвин*), the velocity of the glide was reconstructed by interpolation between the preceding H and the following L target. Local microprosodic fluctuations in pitch tracks due to immediately adjacent obstruents were excluded from the annotations of turning points (e.g., Hanson, 2009).

#### 2.5 Acoustic Measurements

All acoustic measurements (time and frequency) were taken from created EMU labels, using the 2.15.1 version of R and *emu* library (Harrington, 2010). Following measures were calculated:

- velocity  $V$  of the phrase-final pitch change in semitones per second (st/sec):

$$V_{\text{st/sec}} = \frac{12(\log_2 f_a - \log_2 f_b)}{t_b - t_a} \quad (1)$$

In (1), pitch frequency  $f$  (in Hz) and time point  $t$  (in sec) of the start  $a$  and end  $b$  of a (reconstructed or actual, see section 2.4) pitch fall allow for positive values to be calculated. In contrast, negative values are indicative of a pitch rise.

- scaling  $S$  of high or low f0-targets ( $f_b$  in Hz) with respect to the speaker-specific mean f0 ( $f_a$  in Hz) in semitones (st):

$$S_{st} = 12(\log_2 f_b - \log_2 f_a) \quad (2)$$

According to the semitone conversion formula, f0-values below speaker-specific mean are negative while the values above the mean are positive. This transformation of raw f0-values effectively normalized for speaker-specific variance in the dataset and represented measured pitch targets in units of an auditory scale.

- temporal synchronization  $T$  of f0-targets ( $t_{f0}$ ) with the accented vowel  $A$  as a proportion of the vowel duration:

$$T = \frac{t_{f0} - A_{\text{onset}}}{A_{\text{offset}} - A_{\text{onset}}} \quad (3)$$

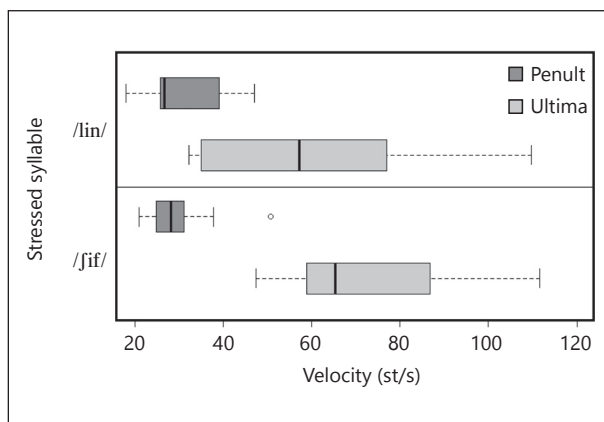
Here, the onset of the accented vowel ( $A_{\text{onset}}$  in sec) equals 0 whereas its offset ( $A_{\text{offset}}$  in sec) equals 1. That is, the formula creates a linear normalization of temporal data where output values below 0 or above 1 indicate that a pitch target is localized before or after the accented vowel, respectively (Silverman and Pierrehumbert, 1990). The vowel was chosen as the unit of normalization to maintain comparability of the time scale across all experimental words, regardless of the segmental composition of their stressed syllable.

### 3 Results

The sections below report the results of acoustic measurements taken from productions by the twenty speakers of Russian and German. Each section concentrates on the following three measurements, seeking to answer the corresponding research questions (cf. 1.3):

- (1) *Velocity of the phrase-final fall*: Was there an increase of velocity if the nuclear accent was located in ultimas in comparison to penults? Did the velocity increase in stressed syllables with obstruents in comparison to sonorants?
- (2) *Scaling of high and low pitch targets* involved in the phrase-final fall: Was the phrase-final L truncated? Was the preceding H perhaps undershot?
- (3) *Alignment of the pitch targets* related to the associated tone(s) of the corresponding pitch accent: Did speakers adjust pitch targets to accommodate for an increased time pressure from the upcoming phrase boundary?

Statistical analyses were conducted in R (version 3.1.0). Factorial analyses employed repeated measures univariate ANOVA. Separate analyses were conducted for each acoustic measurement of interest, i.e. velocity, scaling, alignment each served as a dependent variable. Datasets used for the factorial analyses contained participant-specific means across three repetitions (i.e. 16 values per speaker). The effect of two fixed factors, *stress placement* (ultima, penult) and *consonant type* (obstruent, sonorant), and their interaction were tested. Partial eta squared  $\eta^2$  was calculated to estimate the effect size;  $\eta^2$  indicates what proportion of the variance in the dependent variable is explained alone by the factor in question; and varies from 0 to 1 (the higher  $\eta^2$ , the higher the explanatory power of the factor, the stronger the effect). Where necessary, pairwise comparisons using paired t-test and the Bonferroni adjustment of alpha levels of 0.0125 (0.05/4) subsequently helped to uncover the relevant contrasts and to check for significant differences beyond main effects. Planned comparisons included: (1) penultimate syllables with obstruents vs. sonorants; (2) ultimate syllables with obstruents vs. sonorants; (3) penultimate vs. ultimate stress in syllables with obstruents; (4) penultimate vs. ultimate stress in syllables with sonorants. We start the presentation of the results with medial peaks since



**Fig. 4.** Velocity of phrase-final pitch falls in medial peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress (penult, ultima) is indicated by the two shades of grey (dark, light).

they have the smallest amount of tonal information to be transmitted (H, L), followed by early peaks (H, L, L) and late peaks with the largest amount of tonal information (L, H, L).

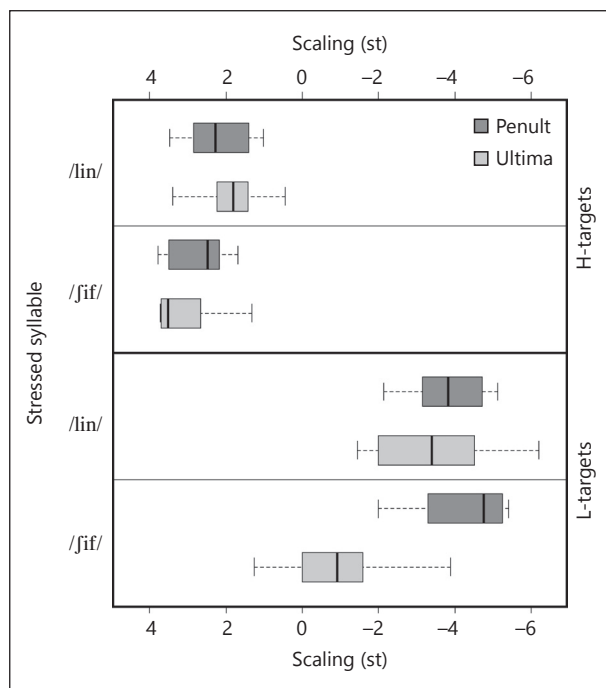
### 3.1 Medial Peaks (H\* L%)

#### 3.1.1 Medial Peaks in Russian

The results for the three acoustic parameters of Russian medial peaks are displayed in figures 4–6. In the boxplots here and below, the dark vertical lines represent the median. Each box includes 50% of the data (i.e. the values lying between the 25th and 75th percentiles). The whiskers mark the most extreme data points outside of the interquartile range. Occasional dots show individual outliers (if there are any).

First analysis showed that the velocity of phrase-final pitch fall in Russian (fig. 4) was affected by an interplay of the two experimental manipulations, *consonant type* and *stress placement* ( $F(1,9) = 6.6$ ,  $\eta^2 = 0.42$ ,  $p < 0.05$ ). Confirming the key observation from figure 4, we found that syllables with ultimate stress measured higher velocities than those with penultimate stress, though the magnitude of the effect was specific to each consonant type: 59 vs. 30 st/sec in syllables with sonorants ( $t(9) = 4.8$ ,  $p < 0.01$ ) and 74 vs. 30 st/sec in syllables with obstruents ( $t(9) = 9.4$ ,  $p < 0.001$ ). The interaction was further indicative of higher pitch velocities measured in the ultimate /fif/ than in the ultimate /lin/ (74 st/s vs. 59 st/s, respectively), though the effect was merely trending toward significance at the set alpha level ( $t(9) = 2.7$ ,  $p = 0.027$ ). No difference was found in the penultimate stress position (30 st/s in both words). These findings are in line with the interpretation of compression and suggest that a substantial increase in the velocity of the phrase-final fall was predominantly triggered by the absence of post-nuclear voicing, and only marginally by a reduced amount of voicing within the phrase-final nuclear syllable itself.

The scaling of H-target (top panel of figure 5) was influenced exclusively by the consonant type of the stressed syllable ( $F(1,9) = 17.5$ ,  $\eta^2 = 0.66$ ,  $p < 0.01$ ). Regardless of stress placement, /fif/ measured 1 st higher targets than /lin/. The finding did not corroborate the predictions of the target truncation model put forward in section 1.3, and deserved an independent explanation (see Sect. 4.1). Overall, H-targets of medial peaks



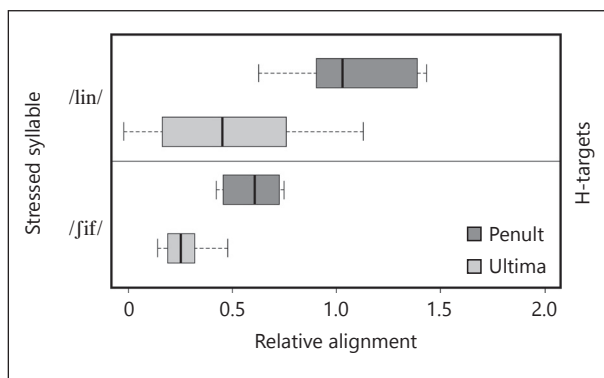
**Fig. 5.** Scaling of low (bottom panel) and high (top panel) pitch targets constituting the phrase-final fall in medial peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

(corresponding to the starred H- tone of the H\* accent) seemed to be relatively stable in frequency.

In contrast, both experimental manipulations significantly influenced the L-target scaling (corresponding to the L% boundary tone, see the bottom panel of figure 5). The interaction of *consonant type* and *stress position* ( $F(1,9) = 24.4$ ,  $\eta^2 = 0.72$ ,  $p < 0.001$ ) demonstrated that the low target in the ultimate /jif/ was scaled significantly higher than in the other target words, i.e. 4 st higher than in the penultimate /jif/ ( $t(9) = 5.0$ ,  $p < 0.001$ ); and 2 st higher than in the ultimate /lin/ ( $t(9) = 4.7$ ,  $p < 0.01$ ). The penultimate /jif/ and /lin/ did not differ significantly from each other, and the 1 st-difference between the penultimate and ultimate /lin/ was very subtle and did not reach significance at the set alpha-level ( $t(9) = 2.7$ ,  $p = 0.024$ ). These results support the view of a continuous L-tone truncation due to a decreasing amount of voicing available in phrase-final positions. The pitch trough of 4.5 st below the speaker's mean is slightly undershot in fully voiced ultimate-stress syllables like /lin/ but strongly undershot in ultimate-stress syllables with obstruents like /jif/. Note though that even in the latter cases, there was still a considerable phrase-final f0-fall of approximately 4 st.

Similarly, the alignment of H (fig. 6) was also significantly influenced by an interplay of the two experimental manipulations: the significant interaction of *stress position* and *consonant type* ( $F(1,9) = 10.4$ ,  $\eta^2 = 0.54$ ,  $p < 0.05$ ) indicated that the magnitude of the temporal re-alignment from late in the penultimate to early/earlier in the ultimate stress syllable depended on the voicing of the corresponding syllable, with a substantial change from 1.2 to 0.5 in the sonorant syllables ( $t(9) = 6.8$ ,  $p < 0.001$ ).





**Fig. 6.** Relative alignment of the high pitch target in medial peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

and a more subtle change from 0.6 to 0.3 in the obstruent syllables ( $t(9) = 9.1$ ,  $p < 0.001$ ). While the 0.6 difference between the means of the penultimate /lin/ and /jif/ reached significance ( $t(9) = 4.1$ ,  $p < 0.01$ ), the 0.2 difference between the ultimate /lin/ and /jif/ syllables did not ( $t(9) = 1.8$ ,  $p = 0.099$ ). The maximal difference in alignment measured 0.9, from the earliest placement of H in *Rashif* (0.3) to the latest placement in *Kalinkin* (1.2). These results can be easily reconciled with the predictions of the alignment model discussed in section 1.3, with the adjustment being sensitive primarily to the amount of postnuclear voicing, and less so to the amount of voicing in the nuclear syllable itself.

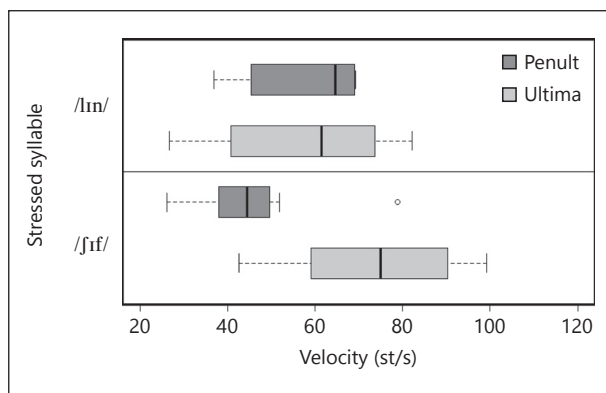
In summary, Russian speakers investigated in this study applied systematic adjustments to all three f0-parameters to produce the f0-patterns necessary for H\* L% in varied amounts of voicing. The adjustments did not only include a simultaneous implementation of L-tone truncation and compression of the phrase-final fall, but also a re-alignment of the associated high tone.

### 3.1.2 Medial Peaks in German

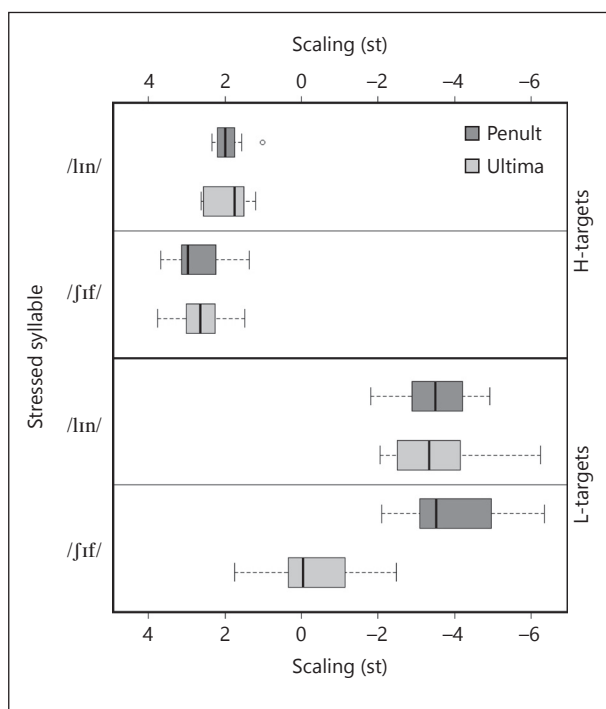
Measurements of velocity, scaling and alignment parameters obtained for German are given in figures 7–9.

First of all, the velocity of the pitch fall (fig. 7) was influenced by the interaction of *consonant type* and *stress placement* ( $F(1,9) = 15.9$ ,  $\eta^2 = 0.64$ ,  $p < 0.01$ ). The ultimate /jif/ measured the highest velocity in these data (80 st/s), significantly different from the penultimate /jif/ (45 st/s;  $t(9) = 3.4$ ,  $p < 0.01$ ) and marginally diverging from the ultimate /lin/ (57 st/s;  $t(9) = 2.2$ ,  $p = 0.055$ ). No other planned comparisons were significant at the set alpha level. As in Russian data presented in section 3.1.1, these findings are in line with a compression account discussed in section 1.3, but in contrast to Russian where the velocity was highly sensitive to a variable amount of post-accentual phrase-final voicing, compression in these German data were observed exclusively in phrase-final stress syllables with obstruents, i.e. under extreme time pressure conditions.

As in Russian, the scaling of H-targets (top panel in figure 8) was affected only by the type of consonant in the accented syllable ( $F(1,9) = 29.8$ ,  $\eta^2 = 0.77$ ,  $p < 0.001$ ). There was a very similar 1 st-increase of peak height measured in syllables with obstruents as compared to those with sonorants, unpredicted by our truncation model.

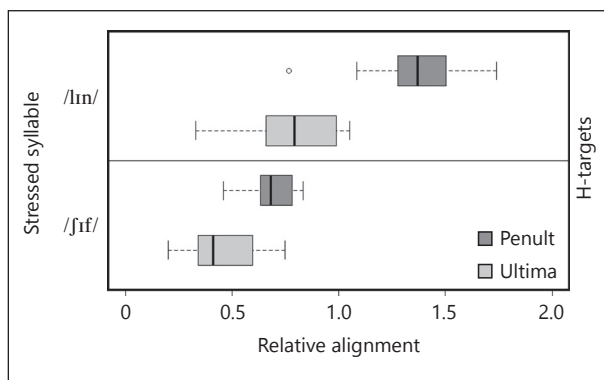


**Fig. 7.** Velocity of phrase-final pitch falls in medial peaks produced by the speakers of German in four target word structures. The position of the syllable carrying lexical stress (ultima, penult) is indicated by the two shades of grey.



**Fig. 8.** Scaling of low (bottom panel) and high (top panel) targets constituting the phrase-final fall in medial peaks produced by the German subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

And again similar to the Russian results reported above, we found a significant interaction of *consonant type* and *stress placement* in the German L-scaling data ( $F(1,9) = 14.3$ ,  $\eta^2 = 0.61$ ,  $p < 0.01$ ). As can be seen in the bottom panel in figure 8, this effect was driven exclusively by the realization of L in the ultimate /fɪf/, which was significantly different from both penultimate /fɪf/ ( $t(9) = 4.7$ ,  $p < 0.01$ ) and ultimate /lɪn/ ( $t(9) = 4.4$ ,  $p < 0.01$ ). None of the remaining planned comparisons turned out significant, meaning that in contrast to Russian, truncation in German occurs only



**Fig. 9.** Relative alignment of the high pitch target in medial peaks produced by the German subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

in cases of extremely limited voicing as in phrase-final nuclei containing only a phonemically short vowel and no other voicing to carry pitch information. However, a 3 st-undershoot of an approximately –3.5 st low target and an average peak height of 2.5 st means that even in the words with little voicing, there is still a pitch fall of 3 st in magnitude to demarcate L%.

Alignment of H was also affected by all experimental manipulations (fig. 9). The significant interaction of *consonant type* and *stress position* ( $F(1,9) = 25.2$ ,  $\eta^2 = 0.50$ ,  $p < 0.001$ ) indicated that details of the alignment were specific to each target word. Words with penultimate stress syllables containing sonorants showed the right-most alignment of 1.3 and differed by approximately 0.5–0.6 time units from both /ln/ in ultimate stress positions ( $t(9) = 7.9$ ,  $p < 0.001$ ) and /ɪf/ in penultimate stress positions ( $t(9) = 9.8$ ,  $p < 0.001$ ). In contrast, words with ultimate stress syllables containing obstruents showed the left-most alignment of 0.5, significantly different from both /ɪf/ in penultimate stress positions (0.7;  $t(9) = 4.0$ ,  $p < 0.01$ ) and /ln/ in ultimate stress positions (0.8;  $t(9) = 6.0$ ,  $p < 0.01$ ). The overall change in alignment amounted to 0.8 (from 0.5 in *Schiff* to 1.3 in *Linner*). Generally speaking, these results demonstrated that the alignment of the H-target was highly sensitive to the segmental environment and the syllabic composition of the nuclear accent and showed a continuous adjustment in line with the alignment model discussed in section 1.3.

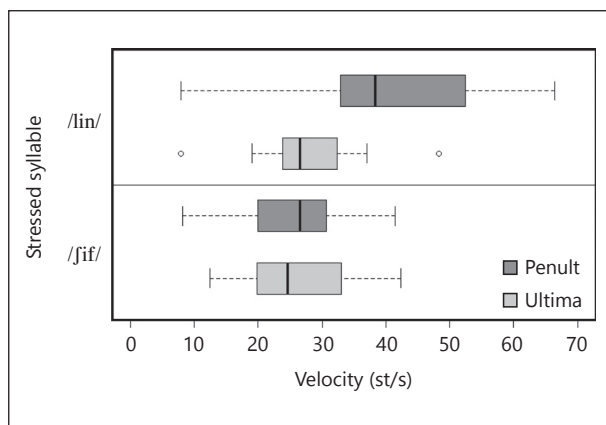
To summarize, German – like Russian – speakers of this study simultaneously applied truncation of L-tone, compression of the phrase-final fall and temporal realignment of the associated tone to accommodate the pitch information relevant to H\* L% when the time span available for voicing was shortened.

### 3.2 Early Peaks (H+L\* L%)

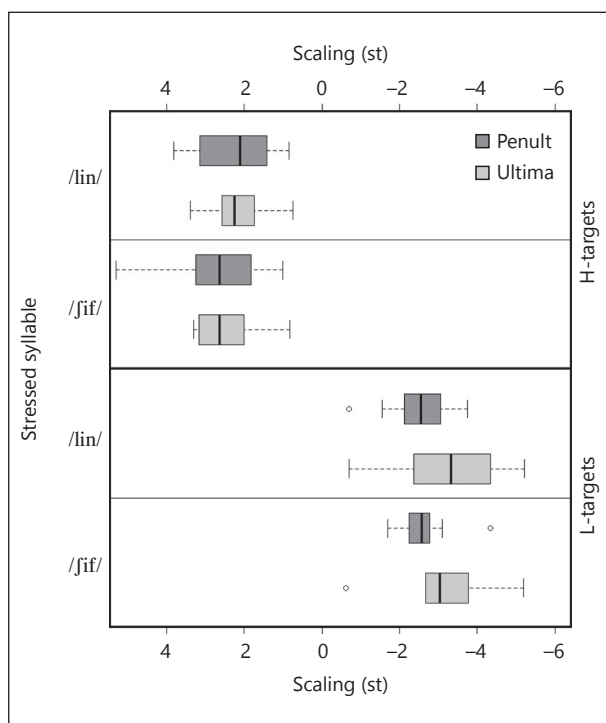
#### 3.2.1 Early Peaks in Russian

Boxplots of velocity, scaling and alignment measured in Russian H+L\* L% are given in figures 10–12.

According to the first statistical analysis, the variance in the data was best explained by an interaction of the two experimental factors, *consonant type* and *stress placement* ( $F(1,9) = 14.2$ ,  $\eta^2 = 0.61$ ,  $p < 0.01$ ). This effect was solely due to higher velocities measured in the penultimate /ln/, significantly different from both the ultimate /ln/ (40 vs. 28 st/s;  $t(9) = 3.6$ ,  $p < 0.01$ ) and the penultimate /ɪf/ (40 vs. 25 st/s;  $t(9) = 3.7$ ,  $p < 0.01$ ).

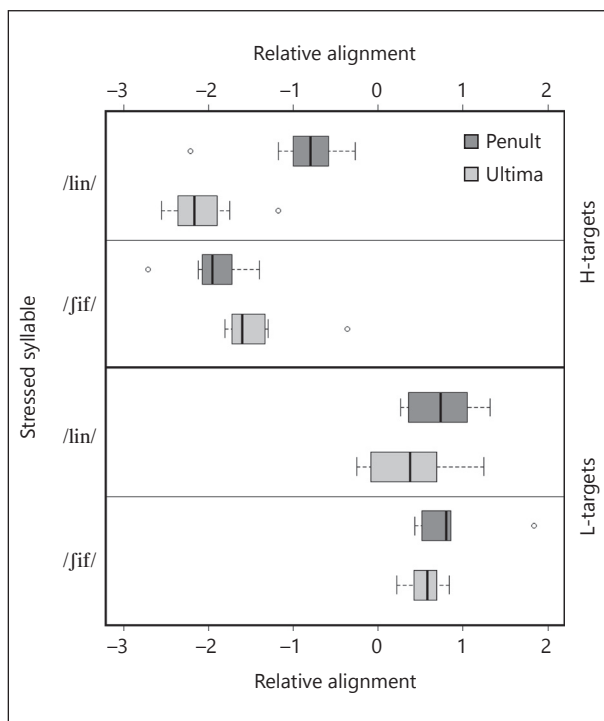


**Fig. 10.** Velocity of phrase-final pitch falls in early peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress (ultima, penult) is indicated by the two shades of grey.



**Fig. 11.** Scaling of low (bottom panel) and high (top panel) pitch targets constituting the phrase-final pitch fall in early peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

All other planned comparisons failed to produce an effect. As can be seen in figure 10, the variability of the velocities measured in the target word *Kalinkin* is quite striking, particularly given a relative consistency of this parameter in all other target words. This finding might be reflective of the individual properties of the target words chosen for investigation: *Kalinkin* is the only token to have no intervening voiceless consonants

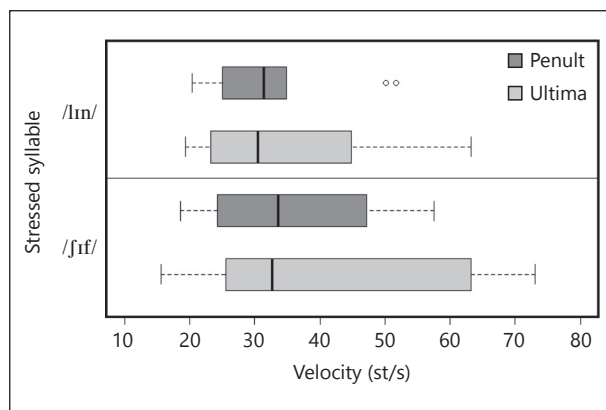


**Fig. 12.** Relative alignment of the high (top panel) and low (bottom panel) targets in early peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

between the prenuclear and nuclear syllables, allowing for more flexibility in the production of the leading tone. Also, since the results for the velocity of a pitch change are not independent of those for alignment and scaling, the overall pattern of early peaks needs to be assessed in order to arrive at an understanding of this finding. In any case, this result is difficult to reconcile with the idea of compression as outlined in section 1.3.

There were no significant results involving the scaling of H (top panel in figure 11). The scaling of L was also quite stable with the sole, weak effect of *stress placement* ( $F(1,9) = 8.1$ ,  $\eta^2 = 0.47$ ,  $p < 0.05$ , cf. bottom panel in figure 11). Unexpectedly, low targets produced in words with ultimate stress were 0.6 st lower than those in words with the penultimate stress. Given this finding, truncation cannot be attested for early peaks in Russian, but rather a combined effect of lowering due to L\* and L% seem to be at work here.

As far as H-alignment was concerned (top panel in figure 12), the analyses revealed a significant interaction of the two predictors ( $F(1,9) = 61.4$ ,  $\eta^2 = 0.87$ ,  $p < 0.001$ ). Although in all target words H appeared before the beginning of the nuclear vowel (with means ranging from  $-1$  to  $-2$ ), the temporal alignment of H differed extremely across experimental conditions: while H in the penultimate /lin/ was closest to the vowel onset, H in the penultimate /jif/ was farthest away from it ( $t(9) = 4.8$ ,  $p < 0.001$ ). And in contrast, ultimate /lin/ had its H-target aligned as early as in the penultimate /jif/ whereas the ultimate /jif/ was aligned closer to the beginning of the nuclear vowel ( $t(9) = 4.1$ ,  $p < 0.01$ ). This alignment pattern is quite different from the re-alignment model that assumes a linear shift to the left (i.e. away from the phrasal



**Fig. 13.** Velocity of phrase-final pitch falls in early peaks produced by the German speakers in four target word structures. The position of the syllable carrying lexical stress (ultima, penult) is indicated by the two shades of grey.

edge) under an increasing time pressure from the right, which should be particularly noticeable in the ultimate /fif/ (see section 1.3). That is, the existing model seems to explain the patterns of these results only poorly. However, the present findings might rather be attributable to some methodological artifacts arising from the chosen normalization procedure since the high target lies far outside the unit of normalization, here the nuclear vowel (cf. Arvaniti et al., 1998). Consequently, we measured the distance between the H-target and the end of voicing in the preceding (i.e. prenuclear) syllable. The new measurement noticeably reduced the variability of H-alignment across the experimental conditions. An additional analysis of variance confirmed that H was consistently aligned 10–20 ms before the end of prenuclear voicing in all target words, regardless of the segmental composition of their stressed syllable.

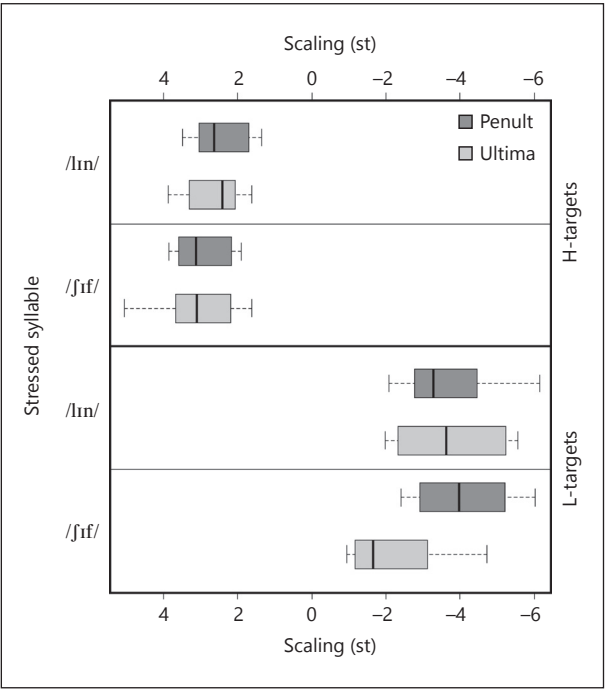
For the alignment of the low target (corresponding to  $L^*$ , see the bottom panel in figure 12), we found a significant main effect of *stress placement* ( $F(1,9) = 24.1$ ,  $\eta^2 = 0.73$ ,  $p < 0.001$ ), with L being aligned later (around 0.9) in words with the penultimate stress and earlier (around 0.4) in words with the ultimate stress. There were no further effects.

In summary,  $H+L^*L\%$  seems to have a robust acoustic representation in Russian. Neither the velocity of the fall nor the frequency of the pitch trough was found to adjust systematically to the decreasing amount of voicing around the nucleus. The only adjustment strategy the Russian speakers employed here was a subtle re-alignment of the low target in keeping with the assumptions of the time pressure model of alignment (see 1.3, see Arvaniti et al., 1998; 2000; Caspers and van Heuven, 1993; D’Imperio, 2001; Prieto and Torreira, 2007). In contrast, the high target was stably ‘anchored’ to the end of voicing preceding the stressed syllable (cf. Arvaniti et al., 1998, 2000).

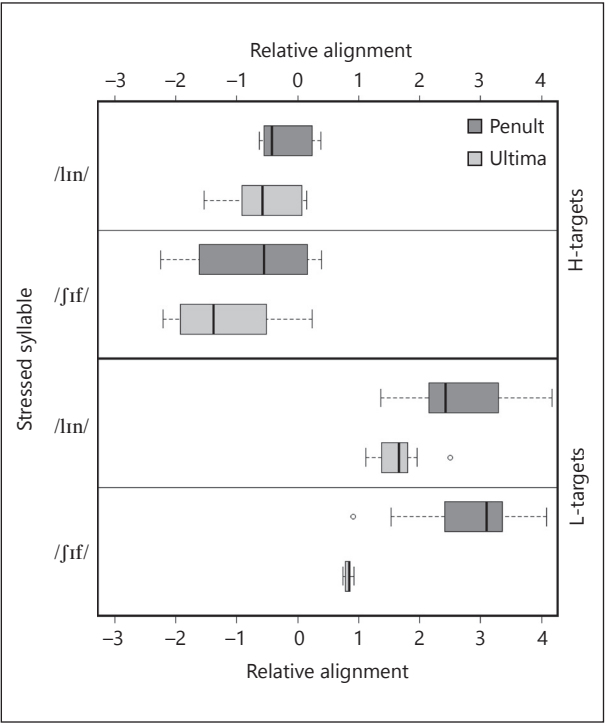
### 3.2.2 Early Peaks in German

Figures 13–15 present boxplots of the measurements resulting from productions of early peaks by the German speakers. As in Russian, not many of the experimental manipulations had a strong impact on the acoustic shape of  $H+L^*L\%$  in German. First of all, the velocity of the fall varied considerably but did not show any significant effect, thus making compression a highly unlikely mechanism here (fig. 13).





**Fig. 14.** Scaling of low (bottom panel) and high (top panel) targets constituting the phrase-final pitch fall in early peaks produced by the ten speakers of German in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.



**Fig. 15.** Relative alignment of the high (top panel) and low (bottom panel) targets in early peaks produced by the ten speakers of German in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

The scaling of H (presented in the top panel of figure 14) was 0.4 st higher if the stressed syllable contained obstruents instead of sonorants ( $F(1,9) = 6.0$ ,  $\eta^2 = 0.40$ ,  $p < 0.05$ ), a result unpredicted by the adjustment model from section 1.3 but already seen above in the findings concerning the realization of medial peaks in both Russian and German (see section 4.1 for a discussion). In contrast, the scaling of L (bottom panel of figure 14) revealed patterns compatible with the idea of truncation. In this analysis, a significant interaction of *stress position* and *consonant type* ( $F(1,9) = 10.6$ ,  $\eta^2 = 0.54$ ,  $p < 0.01$ ) indicated that the penultimate and the ultimate /ln/, and the penultimate /ɟɪf/ all had a comparably scaled L-tone of about  $-4$  st, while this target was undershot by approximately 2 st in the ultimate /ɟɪf/ only (all relevant comparisons were significant at  $t(9) > 2.7$ ,  $p < 0.05$ ).

Two main effects explained the core properties of the H-target alignment (top panel of figure 15). The more influential *stress placement* ( $F(1,9) = 53.7$ ,  $\eta^2 = 0.86$ ,  $p < 0.001$ ) indicated that on average, words with the penultimate stress had a later alignment of high targets than words with the ultimate stress ( $-0.4$  vs.  $-0.8$ , respectively). Note that in both types of words, H was located before the onset of the stressed vowel. A slightly weaker effect of the *consonant type* ( $F(1,9) = 9.6$ ,  $\eta^2 = 0.52$ ,  $p < 0.05$ ) highlighted a similar difference due to this factor influencing the amount of voicing: the H-target was placed further away from the onset of the stressed vowel in syllables with obstruents (around  $-0.9$  time units) than in syllables with sonorants (around  $-0.4$  time units), i.e. H-targets in /ln/ were closer to the vowel onset, yet still preceding the vowel by 0.4 time units. These results are in keeping with the predictions of the alignment model discussed in section 1.3.

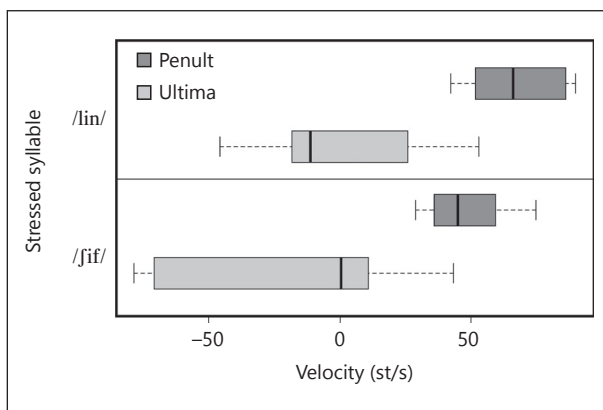
Alignment properties of the low target are shown in the bottom panel of figure 15 (corresponding to  $L^*$ ), and can be best explained by the interaction of *stress placement* and *consonant type* ( $F(1,9) = 9.4$ ,  $\eta^2 = 0.51$ ,  $p < 0.05$ ). Interestingly, the end of the falling movement was reached after the offset of the stressed vowel around 2.6–2.8 time units in all words with penultimate stress (/ln/ and /ɟɪf/ did not differ in this respect). In contrast, the low target was shifted closer to the vowel offset, around 1.7 in the ultimate /ln/, which was significantly earlier than in the penultimate /ln/ ( $t(9) = 4.5$ ,  $p < 0.01$ ). The earliest alignment of 0.8 was found in the ultimate /ɟɪf/, significantly different from both the penultimate /ɟɪf/ ( $t(9) = 6.5$ ,  $p < 0.001$ ) and the ultimate /ln/ ( $t(9) = 7.1$ ,  $p < 0.001$ ). This finding is suggestive of a continuous temporal re-adjustment due to a decreasing amount of voicing until the end of the prosodic phrase, as predicted by the alignment model.

To summarize,  $H+L^*L\%$  in German shows a minimal amount of truncation, occurring only in phrase-final nuclei flanked by voiceless consonants but not elsewhere. No systematic compression effects were found in these data. In contrast, the alignment of both high and low targets was highly sensitive to the segmental and syllabic structure of the target word and showed a gradual shift to the left with an increasing time pressure from the right-hand prosodic boundary, a finding that again is in keeping with the time pressure model of alignment discussed in section 1.3 (Arvaniti et al., 1998, 2000; Caspers and van Heuven, 1993; D’Imperio, 2001).

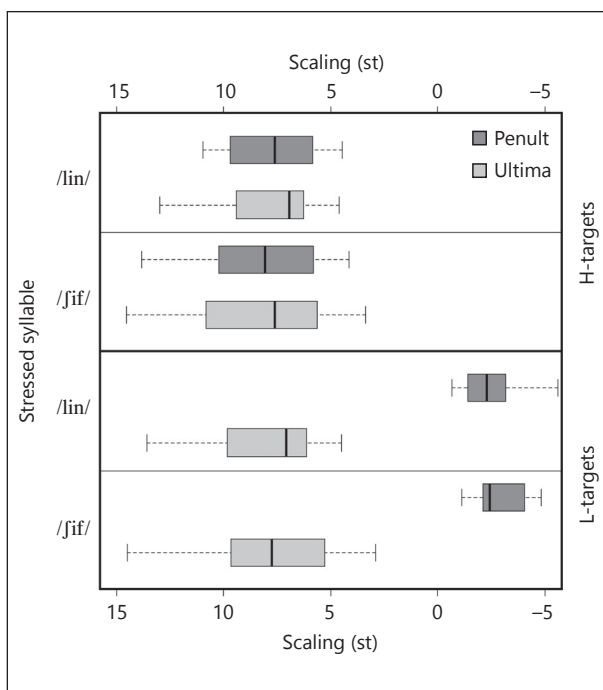
### 3.3 Late Peaks ( $L^*+HL\%$ )

#### 3.3.1 Late Peaks in Russian

The boxplots in figures 16–18 display the measurements of velocity, scaling and alignment in late peaks produced by the Russian speakers.

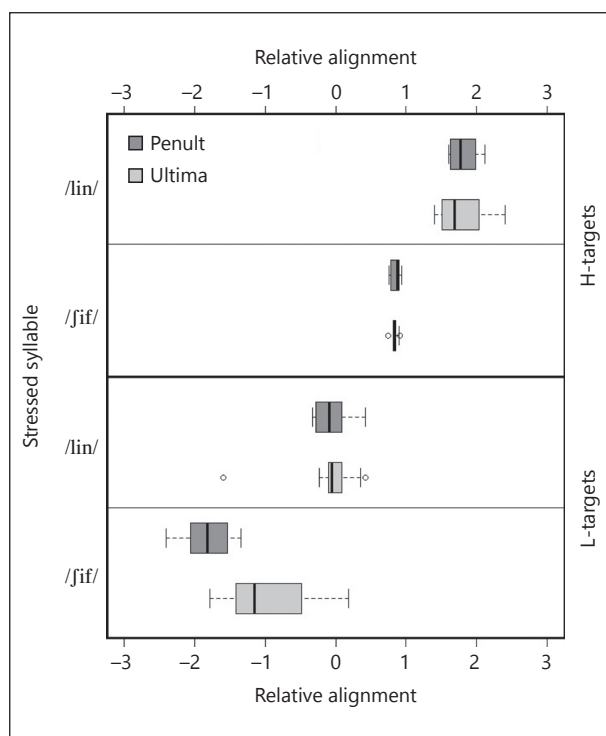


**Fig. 16.** Velocity of phrase-final pitch changes in late peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress (ultima, penult) is indicated by the two shades of grey.



**Fig. 17.** Scaling of low (bottom panel) and high (top panel) targets constituting the phrase-final pitch pattern in late peaks produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

The first statistical analysis was conducted for the dependent variable velocity of the fall (fig. 16), and produced two significant main effects, *consonant type* ( $F(1,9) = 5.3$ ,  $\eta^2 = 0.37$ ,  $p < 0.05$ ) and *stress placement* ( $F(1,9) = 23.2$ ,  $\eta^2 = 0.72$ ,  $p < 0.001$ ) but no interaction. According to this analysis, the velocity of the phrase-final pitch fall was primarily affected by the placement of the stressed syllable. In the penultimate stress, the velocity of the fall amounted to 57 st/s on average. However, no fall but a slight rise of about 10 st/s was found in words with ultimate stress. Regarding the main effect



**Fig. 18.** Relative alignment of the high (top panel) and low (bottom panel) targets in L\*+H pitch accents produced by the Russian subjects in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

of *consonant type* ( $F(1,9) = 5.3$ ,  $\eta^2 = 0.37$ ,  $p < 0.05$ ), stressed syllables with obstruents had on average a 15 st/s lower velocity of the fall than syllables with sonorants (i.e. 49 st/s vs. 65 st/s). These patterns cannot be interpreted as instances of compression but rather showcase a different strategy for dealing with the time pressure arising from a limited amount of phrase-final voicing. These results are suggestive of pitch targets being aligned independently of each other and of the resulting velocity, in order to accommodate the relevant pitch height information in the voicing available around the stressed syllable (cf. alignment results below).

The aforementioned finding was further supported by the statistical output for the scaling of the phrase-final L-target at the two levels of *stress placement* ( $F(1,9) = 73.1$ ,  $\eta^2 = 0.89$ ,  $p < 0.001$ , cf. bottom panel of figure 17). Phrase-final pitch was significantly different in words with the ultimate stress which ended high, in contrast to words with the penultimate stress which ended low. Subsequent t-tests revealed that the phrase-final high pitch in the ultimate /lin/ and /jif/ did not significantly differ from the preceding H-target. That is, the phrase-final fall was completely abandoned if the nuclear syllable carrying L\*+H L% ended a prosodic phrase.

There were no significant effects for H-scaling (top panel in figure 17), neither was the scaling of the preceding associated L-tone affected by experimental manipulations (the latter is not plotted in figure 17).

As far as the alignment of high targets in L\*+H was concerned (top panel in figure 18), only the *consonant type* ( $F(1,9) = 136.1$ ,  $\eta^2 = 0.94$ ,  $p < 0.001$ ) played a significant

role, with a large alignment difference of 1.0 between syllables with obstruents (generally aligned with the stressed vowel around 0.8) and syllables with sonorants (generally aligned around 1.8, i.e. after the stressed vowel). Strikingly, there was little variability in the alignment of high targets with the syllables containing obstruents, quite possibly driven by the limited duration of voicing in those syllables and the need for a consistent late alignment.

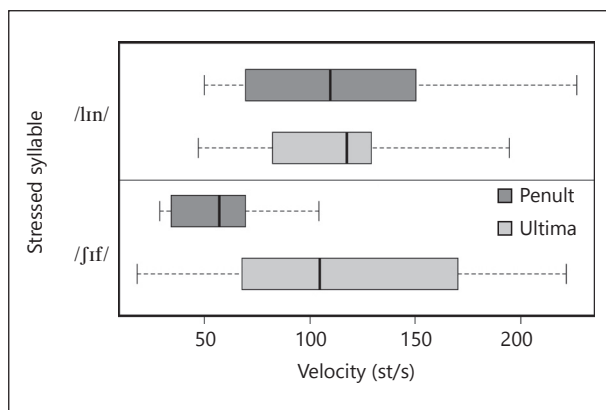
The temporal alignment of the low target corresponding to the associated tone in L\*+H (shown in the bottom panel of figure 18) varied significantly due to an interplay of the two experimental manipulations. The significant interaction of *consonant type* and *stress placement* ( $F(1,9) = 9.9$ ,  $\eta^2 = 0.52$ ,  $p < 0.05$ ) was indicative of a differential response of syllables with obstruents vs. sonorants to the implemented variability in the stress placement. All words containing /lin/ showed little difference in the alignment patterns of L ( $t(9) = 0.4$ , n.s.), with all targets being aligned slightly before the onset of the stressed vowel at -0.1. In contrast, the two sets of words containing /jif/ had a pronounced difference of 0.8 alignment units ( $t(9) = 3.2$ ,  $p = 0.011$ ) between words with the penultimate stress (-1.8) and the ultimate stress (-1.0). As in the early peak results discussed in section 3.2.1, this alignment pattern was rather unexpected in the light of the re-alignment model put forward in section 1.3. Suspecting an artifact of the normalization procedure due to the fact that the low target lies outside of the unit of normalization (cf. Arvaniti et al., 1998), we additionally measured (1) the distance between the L-target and the start of the fricative in /jif/-words and (2) the distance between the L-target and the start of the nuclear vowel in /lin/-words. Again, the measurement helped to substantially streamline the variability in the L-target alignment data, with a clear pattern arising for both types of stressed syllables. A subsequently conducted ANOVA confirmed that the L-target was consistently aligned with the voicing available 10 ms before the onset of the nuclear vowel (which was located in the pre-nuclear syllable in words containing /jif/ but in the onset sonorant of the nuclear syllable in words containing /lin/). Accordingly, the significant difference between the two types of /jif/-words reported above might have arisen mainly due to some differences in the timing of nuclear vowels produced in those words, and not due to the experimental manipulations of this study.

In summary, the key feature of Russian late peaks seemed to be a somewhat radical truncation of the phrase-final low targets in phrases with the ultimate nucleus. Again, there was a re-alignment of the high tones in response to the changes in the availability of time and voicing, while the associated low tones were aligned consistently with the voicing preceding the nuclear accented vowel (cf. Arvaniti et al., 1998, 2000).

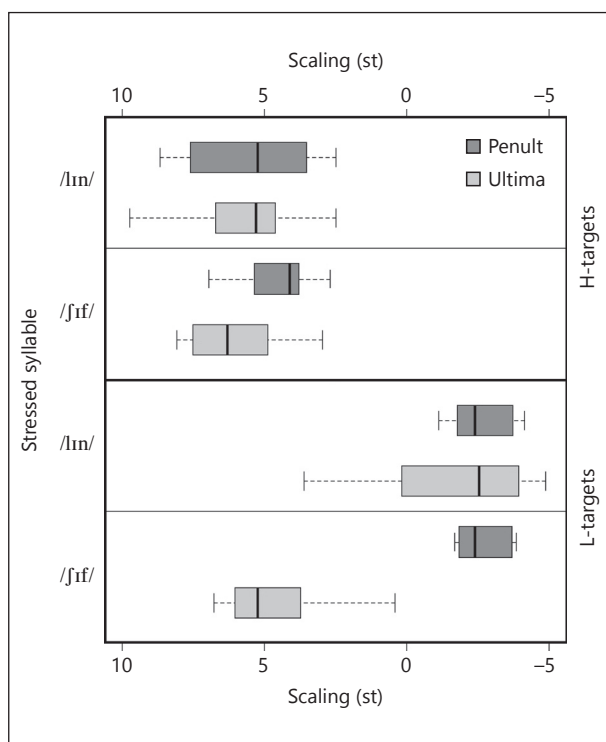
### 3.3.2 Late Peaks in German

The graphs below present the findings for the velocity of the fall (fig. 19), the scaling of the pitch targets (fig. 20) and their alignment (fig. 21), all measured in late peaks produced by the German speakers.

The large variability in the velocity of the phrase-final fall shown in figure 19 did not produce any consistent pattern to be backed up by statistics, and there were no signs of a systematic use of velocity of the fall in terms of compression. Changes in the velocity of the accentual rise (corresponding to L\*+H, not included in the figures) did not show any signs of compression, either. Rather an opposite effect was found: rises in syllables flanked by obstruents were approximately 20 st/sec slower than in syllables with sonorants ( $F(1,9) = 10.5$ ,  $\eta^2 = 0.52$ ,  $p < 0.01$ ).



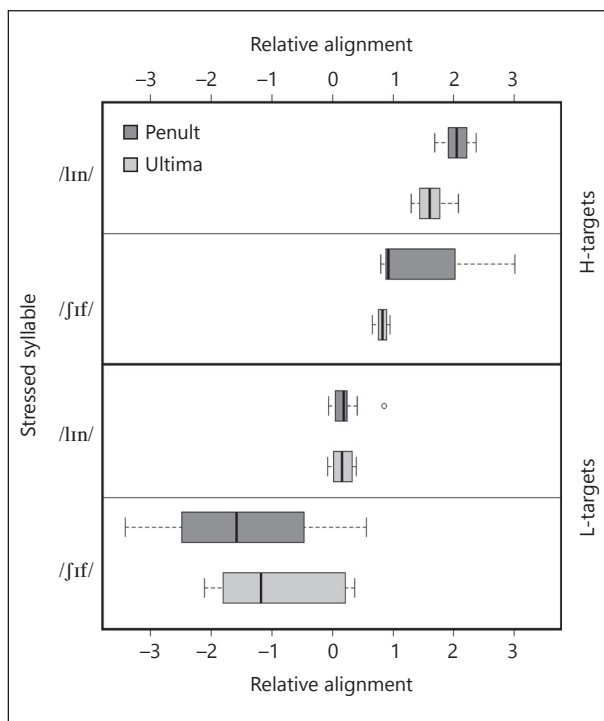
**Fig. 19.** Velocity of phrase-final pitch falls in late peaks produced by the German subjects in four target word structures. The position of the syllable carrying lexical stress (ultima, penult) is indicated by the two shades of grey.



**Fig. 20.** Scaling of low (bottom panel) and high (top panel) targets constituting the phrase-final fall in late peaks produced by the German speakers in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

The scaling of H (shown in the top panel of figure 20) was slightly influenced by the interaction of *stress position* and *consonant type* ( $F(1,9) = 9.8$ ,  $\eta^2 = 0.52$ ,  $p < 0.05$ ). This effect was attributable to the H-target being 1 st lower in the penultimate /ɪf/ than in the ultimate /ɪf/ ( $t(9) = 4.5$ ,  $p < 0.01$ ) while the difference between the two /ln/-words was not significant ( $t(9) = 0.1$ , n.s.).





**Fig. 21.** Relative alignment of the trailing H (top panel) and low L\* (bottom panel) targets in late peaks produced by the German speakers in four target word structures. The position of the syllable carrying lexical stress is indicated by the two shades of grey.

The scaling of the phrase-final pitch trough (bottom panel in figure 20) was best explained by an interplay of the two experimental factors, *consonant type* and *stress position* ( $F(1,9) = 32.3$ ,  $\eta^2 = 0.78$ ,  $p < 0.001$ ), attributable mainly to the ultimate /ʃɪf/ being an outlier, as expected by the truncation model in section 1.3. As revealed by a series of t-tests, the patterns of L%-scaling were comparable (i.e. not statistically different) for all target words but for those with the ultimate /ʃɪf/ whose L-scaling was approximately 6.5 st higher than in the ultimate /lɪn/ ( $t(9) = 5.8$ ,  $p < 0.001$ ) or in the penultimate /ʃɪf/ ( $t(9) = 9.0$ ,  $p < 0.001$ ). Unlike in Russian late peaks though, there was still a fall of about 1.5 st in phrase-final positions of German late peaks, suggesting that truncation of L\*+H L% is categorically different in these two languages. In contrast, the scaling of the associated L\*-tone (not plotted) was not affected by any of the experimental manipulations.

Both experimental manipulations produced a significant main effect on the alignment properties of high targets (top panel in figure 21). Varying *consonant type* ( $F(1,9) = 35.7$ ,  $\eta^2 = 0.80$ ,  $p < 0.001$ ) caused an alignment difference of approximately 0.7 time units between words containing sonorants (1.1) and in words containing obstruents (1.8). In both cases, however, the H-targets were located relatively late with respect to the stressed vowel. The manipulation of *stress placement* produced a slightly weaker effect ( $F(1,9) = 11.2$ ,  $\eta^2 = 0.56$ ,  $p < 0.01$ ), though in line with the assumptions of the time pressure model in section 1.3: H-alignment in words with the ultimate stress was about 0.5 time units earlier than in words with the penultimate stress (1.7 vs. 1.2, respectively).

In contrast, the alignment of the associated L\*-tone target was most strongly influenced by the interaction of *stress placement* and *consonant type* ( $F(1,9) = 33.1$ ,  $\eta^2 = 0.79$ ,  $p < 0.001$ , cf. bottom panel in figure 21). This finding was shaped mainly by the alignment differences observed among the two /ʃɪf/-words with the penultimate vs. ultimate stress placement since the slight difference between the two /lɪn/-words was not significant. In words containing /lɪn/, L was aligned early within the nuclear vowel at around 0.2 time units. In words containing /ʃɪf/, we observed a much earlier alignment of L in the penultimate (−1.8) than in the ultimate (−0.9) stress ( $t(9) = 4.4$ ,  $p < 0.01$ ), similar to the effect reported for late peaks in Russian. Once again, given that this result was at odds with the alignment model discussed in section 1.3, we subsequently checked for an alternative way of understanding these data and used a non-normalized measurement of temporal alignment. However, in contrast to Russian, the resulting alignment patterns were not as clear-cut: the German speakers varied in the alignment of the L target, either producing it outside of the nuclear syllable (in 30% of the words with the penultimate stress and in 43% of the words with the ultimate stress) or within the nuclear accented vowel (70% of the words with the penultimate stress and 57% of the words with the ultimate stress). When aligned before the stressed syllable, L-targets occurred 16 ms before the onset of /ʃ/ in /ʃɪf/. If aligned within the stressed vowel, L-targets were placed 28 ms into the vowel. Given the mean vowel duration of 91 ms in this subset of the data, such target alignment is comparable to the patterns ascertained in the syllables with sonorants (i.e., 0.2–0.3 normalized time units). These two early alignment choices (preceding the stressed syllable vs. early in the stressed vowel) were present both within the same speaker and between speakers. Although the results were not statistically significant, we note that a shift of the target outside of the nuclear vowel occurred predominantly in words with the ultimate stress.

To summarize, the result patterns found for German late peaks showed once again that the target alignment was sensitive to the segmental and syllabic structure of the target words. Truncation of the phrase-final pitch trough was produced by the speakers exclusively in words with the ultimate /ʃɪf/ but not elsewhere, i.e. it occurred exclusively under maximal time pressure.

## 4 Discussion

### 4.1 Cross-Linguistic Comparisons of Pitch-Segment Interactions

The present study investigated pitch-segment interactions in early, medial and late peaks (analyzed as H+L\*, H\* and L\*+H pitch accents with a low boundary tone) in Russian and German. Sets of materials were tested in which the phonological structure of phrase-final, nuclear accented words was varied systematically by shortening the amount of material available for voicing from longer words with a postnuclear syllable (German *Linner*; Russian *Kalinkin*) to extremely short words with nuclei flanked by voiceless consonants and no further syllables following the nucleus (German *Schiff*; Russian *Rashif*). The analyses concentrated on three pitch parameters indicative of compression, truncation and temporal compensation: velocity of the fall, scaling and alignment of low and high targets. The results demonstrated that Russian and German speakers used all of the tested adjustment strategies to accommodate complex pitch movements of rises and falls to a shorter period of voicing available to them in the different types of materials. A comparative summary of the results is given in table 4.

**Table 4.** Comparative overview of the results of the present study with regard to the predictions of the pitch adjustment model presented in section 1.3.

	Early peaks H+L* L%	Medial peaks H* L%	Late peaks L*+H L%
Russian	no compression no truncation stable alignment of H+ re-alignment of L*	compression gradual truncation re-alignment of H*	no compression categorical truncation re-alignment of +H stable alignment of L*
German	no compression gradual truncation re-alignment of H+ re-alignment of L*	compression gradual truncation re-alignment of H	no compression gradual truncation re-alignment of +H two choices of alignment of L*

As can be seen in table 4, medial peaks (H\* L%) evoked a combination of compression, truncation and a temporal re-alignment of pitch targets in both languages, showing that it was possible for all three adjustments to be at play simultaneously. As expected, cross-linguistic differences involved primarily the magnitude of these adjustments. For example, the maximal change in velocity of the fall measured in Russian was about 40 st/sec in contrast to 35 st/sec in German. The maximal amount of truncation was also slightly larger in Russian (4 st) than in German (3 st). While the overall alignment shift was comparable across the two languages and amounted approximately to the duration of an accented vowel (0.8–0.9 times units), H was aligned slightly later in German (from 0.5 in *Schiff* to 1.3 in *Linner*) than in Russian (from 0.3 in *Rashif* to 1.2 in *Kalinkin*).

In contrast to medial peaks, Russian early peaks (H+L\*) showed a fairly constant phonetic implementation in spite of variable segmental environments. No target undershoot or systematic increase in the velocity of the fall could be attested. The high tone was consistently aligned around 10–20 ms before the end of voicing in the prenuclear syllable, and only the low tone was shifted from the vowel offset closer to the vowel onset in words without a postnuclear syllable (*Rashif*, *Zhaklin*). Early peaks produced by the German speakers were similar – the alignment of the low target was earlier when the time pressure due to the upcoming phrasal boundary increased, moving the low target into the accented vowel. But once again, cross-linguistic comparisons revealed an earlier target alignment in Russian than in German: while the Russian speakers consistently aligned their L\*-targets with the accented vowel (ranging from as early as 0.3 in *Rashif* to no later than 0.7 in *Kalinkin*), the German speakers produced pitch falls which mostly ended after the stressed vowel (cf. 2.6 in *Linner*) or late in a stressed vowel under maximal time pressure (cf. 0.8 in *Schiff*). Moreover, the German speakers also gradually shifted the high targets further away from the accented vowel and employed a small amount of truncation in words with the maximal time pressure (i.e. in *Schiff*), thus showing a larger scope of pitch modifications in response to the segmental variability than the Russian speakers. Interestingly, Russian words with the ultimate stress measured lower pitch troughs than words with the penultimate stress, contrary to the idea of truncation supported by the German data. Rather, this result points toward a combined effect of pitch lowering due to the presence of two low tones, L\* and L%.

In both Russian and German, the main tonal adjustment in late peaks was truncation, accompanied by a re-alignment of high and low tones. However, truncation surfaced very differently in the two languages. The difference appears to be best described as the one between a gradual target undershoot (in German) and a categorical abandonment of an underlying boundary tone (in Russian). Indeed, an undershoot of 6.5 st in German late peaks produced in the phrase-final *Schiff* (but not *Schiffer*, *Linn* or *Linner*) was stronger than in medial (3 st) or early peaks (2 st), and yet it did not lead to a complete truncation of the phrase-final fall. In Russian, on the other hand, no evidence of a fall was found either in *Zhaklin* or in *Rashif*, i.e. in words with the ultimate stress, regardless of their segmental composition. This result was qualitatively different from a gradual truncation observed in Russian medial peaks, and in the German data.

Overall, the core difference between Russian and German seems to lie not only in the phonetic detail of the three adjustment strategies (truncation, compression and re-alignment), but also in their structurally different responses to the time pressure factors tested here, availability of voicing within vs. after the nuclear accented syllable. Russian seemed to be particularly sensitive to the absence or presence of a postnuclear syllable, with all of the adjustment strategies having a stronger effect in phrase-final nuclei and a weaker effect in syllables flanked by voiceless consonants. In contrast, German was most sensitive to the presence or absence of voicing in the nuclear syllable, with the phrase-final /ʃɪf/ evoking all of the adjustments, with little or no effect in all other segmental environments. Notably, the core similarity of the two languages seemed to lie in their universal application of tonal re-alignment across variable segmental environments. The temporal synchronization of both high and low targets was highly sensitive to the segmental structure and showed a gradual shift to the left with an increasing time pressure from the right-hand prosodic boundary, a finding that again supported the time pressure model of alignment discussed in section 1.3 (e.g., Arvaniti et al., 1998, 2000; Caspers and van Heuven, 1993; D'Imperio, 2001; Prieto et al., 1995; Prieto, 2005; Schepman et al., 2006; Silverman and Pierrehumbert, 1990; Steele, 1986).

With respect to the main aim of this study outlined in section 1.3, the results provided evidence that the two languages classified as 'truncating' (Grabe, 1998; Igarashi, 2002; Odé, 2005; Uhmman, 1991) differed greatly in the amount and type of truncation they utilized, yet unanimously employing further pitch adjustments to transfer the relevant pitch information. The accentual fall in H+L\* was not truncated in Russian and only slightly undershot in German, whereas boundary-related pitch fall following H\* and L\*+H showed truncation effects in both languages, suggesting that phrase-final pitch modifications due to time pressure are sensitive to the phonological composition of the tonal string. According to these results, the previous finding that '*German truncates falls and compresses rises*' (Grabe, 1998: 140) does not seem to apply to all falls and rises, in general. More specifically, pitch targets corresponding to low boundary tones in German were strongly truncated after H\* and L\*+H (3 to 6.5 st undershoot), while the falls corresponding to H+L\* pitch accents were undershot by merely 2 st, by and large preserving the low target of L\*. Moreover, no compression was found in the accentual rise of German L\*+H. Instead, the targets were re-aligned utilizing the temporal compensation strategy.

At the first sight, the German truncation patterns appear to be reconcilable with the idea of a gradual target undershoot as a function of pitch peak alignment regardless of its phonological composition; the closer an accentual peak to the phrasal boundary, the stronger the effect. However, a correlation analysis did not lend substance to the idea of

a systematic relationship between these two acoustic parameters in the overall dataset ( $R^2 = -0.14$ , n.s.), suggesting that truncation in German cannot be modeled linearly but applies differentially to the three pitch categories. Indeed, further correlation analyses revealed that the two parameters were completely independent of each other in early peaks ( $R^2 = -0.01$ , n.s.), but had a weak relationship in medial peaks ( $R^2 = 0.40$ ,  $p < 0.05$ ), and a slightly stronger one in late peaks ( $R^2 = -0.52$ ,  $p < 0.001$ ). In Russian, on the other hand, none of these correlations turned out to be significant, further supporting the above interpretation that truncation is not a homogenous phenomenon but has qualitatively different implementations both cross-linguistically and across different phonological categories within one language.

Why did we find the unpredicted cross-linguistic effect of a slight H-tone upscaling in syllables flanked by obstruents carrying medial peaks? It was also present in early peaks in German but absent in Russian. At first, the effect was suggestive of a microprosodic influence of surrounding voiceless consonants (e.g., Hanson, 2009). However, being (a) not very close to a voiceless consonant and (b) of a considerable magnitude (1 st), the non-local upscaling may be too prominent for a purely microprosodically driven effect. Perhaps, the exact location of high turning points might have been obscured by the intermittent voicelessness in some cases, thus making the upscaling an artefact of the measurement rather than a true instantiation of pitch-segment interactions. Future research will benefit from more fine-grained phonetic measures that take into account different stages of pitch production, not only the turning points but also high and low plateaus (fig. 3; Xu and Sun, 2002).

To summarize, the results of the present study suggest that the classification into ‘truncating’ vs. ‘compressing’ languages can be rather misleading, given that both strategies can apply within a language either simultaneously or selectively, depending on the pitch category. Since intonational phonologies of languages vary greatly (cf. Jun, 2005, 2014), it is difficult to devise a diagnostic tool for future language classification, making this endeavor methodologically questionable and unreliable. Moreover, linguistic affiliation with the truncating group obscures the fact that truncation is not a unified phenomenon and can have different implementations, ranging from a gradual cut-off in pitch frequency to a categorical abandonment of a boundary tone.

#### 4.2 Theoretical Implications of the Results

By and large, the results presented here contribute to the steadily growing body of evidence supporting the key premise of autosegmental-metrical phonology of intonation that tonal units align with segmental strings independently of each other (e.g., Arvaniti et al., 1998; Ladd, 1996/2008; Prieto, 2005; Prieto et al., 1995). However, some proposals also suggest that a succession of two or more tones may differ in their internal structure constituting either a *melodic unit* or a *tonal sequence* (Yip, 1989; Grice, 1995a, b). Tones of a melodic unit are organized with respect to each other and associate to a tone-bearing unit (t.b.u.) as a holistic gesture (Xu, 1998). Consequently, they tend to preserve their temporal and frequency relationship under varying phonetic conditions. In contrast, tones appearing as a sequence establish an association to a t.b.u. relatively independently of each other. Therefore, they are scaled and aligned only with respect to the t.b.u., so that their realizations can deviate considerably in time pressure situations. In keeping with this proposal, we might expect that melodic units undergo systematic compression and re-alignment effects of comparable magnitude in contrast to tonal sequences with highly variable degrees of truncation and re-alignment

affecting their tonal constituents. Grice (1995a, b) argues that the underlying structure of tonal representations can differ across categories of a language and cross-linguistically. For example, leading tones ( $\underline{T}+T^*$ ) in English are implemented as sequences while trailing tones ( $T^*+\underline{T}$ ) are defined as melodic units; in contrast, all pitch accents of Palermo Italian should be analyzed as tonal sequences (Grice, 1995b: 200). According to this proposal, our data shows evidence for tonal sequences being more common than melodic units in both Russian and German. Can this logic be applied to medial peaks and interpreted as evidence for their analysis as  $H^*+L$  in both languages? Perhaps not necessarily, since the low boundary tone  $L\%$  was found to be truncated following bitonal, not monotonal pitch accents (e.g. after  $L^*+H$  in both languages and after  $H+L^*$  in German, but not after  $H+L^*$  in Russian). The differences may be related to a tonal saturation of a t.b.u. which is reached after two tones in these two languages. In sum, much empirical work is needed to elaborate on a comprehensive theory of the structure of tonal representations which accommodates – and benefits from – evidence gathered in diverse segmental environments (cf. Barnes et al., 2012).

The results also have implications for the assumption that intonation languages can be classified as either ‘truncating’ or ‘compressing’ (Ladd, 1996/2008). Since truncation and compression are not mutually exclusive and patterns of truncation (gradual vs. categorical) differ both within and across truncating languages, the dichotomy seems to be of little help in understanding the true nature of the universal vs. language-specific pitch-segment interactions. Instead of searching for a way to divide languages into groups, we should perhaps rather ask the more intriguing question how is it possible that languages manage to maintain a functioning system of intonational contrast in segmental environments with variable availability of time and voicing. It might further be helpful to consider pitch-segment interactions in the context of the H&H theory (Lindblom, 1990) which sees the primary goal of a sound linguistic system in its ability to keep the relevant categories perceptually distinct, while also allowing for some reduction of speech production efforts. Accordingly, we would predict a certain tension to exist between the use of truncation (which keeps the pitch production efforts minimal) and compression (which maintains the perceptual distance between categories). It would make sense for any linguistic system to combine the two strategies to achieve an optimal balance in the sense of H&H. Re-alignment appears to constitute an effective pitch production strategy, though potentially posing the danger of a perceptual confusion of early and medial or medial and late peaks. This account would predict an abundance of perceptual compensation effects similar to those found in segmental co-articulation (e.g., Harrington et al., 2008), and deserves some attention in future research.

#### 4.3 Conclusions and Outlook

In conclusion, the present study demonstrated that the binary typology of ‘truncating’ or ‘compressing’ languages is not a useful cross-linguistic taxonomy. Languages classified as truncating can employ different patterns of truncation (categorical in Russian vs. gradual in German late peaks), prohibiting any attempt at a cross-linguistic generalization about the nature of truncation. Moreover, compression and truncation are not two mutually exclusive mechanisms of  $f_0$ -adjustments, and can operate simultaneously while being sensitive to the language-specific constraints involving the phonological status of different tonal events. It seems plausible to expect that similar patterns may exist within ‘compressing’ languages like English (Grabe, 1998; Ladd, 1996/2008), but have remained hidden because the investigators’ gaze has to date mostly been directed



towards productions of H\* L% and L\* H%, leaving other possible intonational choices out of the picture. And finally, the binary typology does not take into account the third, utterly ubiquitous, way to adjust pitch patterns in phrase-final time pressure situations. Re-alignment of high and low targets has been attested in all pitch accents studied here, and affected most of the tonal targets. The cross-linguistic data presented in this study suggest that the phrase-final pitch-segment interactions in the context of a variable time pressure are sensitive to the phonological composition of the tonal string and the status of a particular tonal event (associated vs. boundary tone), and do not apply to falling vs. rising pitch contours across the board, even within one language.

The present study looked exclusively at pitch adjustments under the influence of segmental variability. However, the pitch-segment interaction is not a one-way street, and cases have been reported where tonal languages lengthen vowels to accommodate contour tones (e.g., Mitla Zapotec, Wuyi Chinese, Hause and Gã, Zhang, 2004) or individual speakers of intonation languages make use of vowel lengthening in contexts of compromised time and voicing to avoid truncation (Prieto and Ortega-Llebaría, 2009). Spectral cues of consonants have also been shown to encode some of the variations related to pitch height (Niebuhr, 2012) and may be at play in the context of phrase-final pitch-segment interactions. Similarly, the present study did not pay any attention to the duration of plateaus, leaving an important ingredient of all pitch peaks out of the complex picture (fig. 3; cf. Knight and Nolan, 2006). These are all promising avenues to deepening our understanding of the interplay between segments and pitch, and may be fruitfully exploited in future cross-linguistic research.

## Acknowledgements

This research was carried out at the *Institute of Phonetics and Speech Processing* in Munich and the *Institute of Phonetics and Digital Speech Processing* in Kiel as a partial fulfilment of the requirements for the PhD degree. I would like to acknowledge inspiring discussions with my PhD supervisor Jonathan Harrington and also Christine Mooshammer, Michel Scheffers, Bob Ladd and the *Phonetics and Phonology* research group at the University of Edinburgh, and thank the editorial team and two anonymous reviewers for their help in improving this manuscript.

## References

- Adank P, Hagoort P, Bekkering H (2010): Imitation improves language comprehension. *Psychol Sci* 21:1903–1909.
- Arvaniti A, Ladd DR, Mennen I (1998): Stability of tonal alignment: the case of Greek prenuclear accents. *J Phon* 26:3–25.
- Arvaniti A, Ladd DR, Mennen I (2000): What is a starred tone? Evidence from Greek; in Broe MB, Pierrehumbert JB (eds): *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge, Cambridge University Press, pp 119–131.
- Babel ME (2012): Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J Phon* 40:177–189.
- Bannert R, Bredvad A (1975): Temporal Organisation of Swedish Tonal Accent: The Effect of Vowel Duration. Working Papers 10. Phonetics Laboratory, Department of General Linguistics, Lund University, Sweden.
- Barnes J, Veilleux N, Brugos A, Shattuck-Hufnagel S (2012): Tonal center of gravity: a global approach to tonal implementation in a level-based intonational phonology. *Lab Phonol* 3:337–383.
- Boersma P, Weenink D (1997): Doing phonetics by computer. <http://www.praat.org>.
- Bondarko LV (1998): *Fonetika sovremennogo russkogo jazyka*. St.-Petersburg.
- Bryzgunova EA (1977): *Zvuki i intonacija russkoj reči*. Moskva, Russkij jazyk.
- Bryzgunova EA (1980): *Intonacija. Russkaja grammatika*. Moskva, Akademija Nauk SSSR, pp 96–122.
- Caspers J, van Heuven VJ (1993): Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50:161–171.

- Crespo-Sendra V, Vanrell MM, Prieto P (2010): Information seeking questions and incredulity questions: Gradient or categorical contrast? *Proceedings of Speech Prosody*, Chicago, USA.
- Dufour S, Nguyen N (2013): How much imitation is there in a shadowing task? *Front Psychol* 4:1–7.
- Eriksson Y, Alstermark M (1972): Fundamental frequency correlates of the grave word accent in Swedish: the effect of vowel duration. *Speech Transmission Laboratory. Quarterly Progress and Status Report* 2–3, KTH, Sweden.
- Goldinger SD (1998): Echoes of echoes? An episodic theory of lexical access. *Psychol Rev* 105:251–279.
- Gordon M (2001): A typology of contour tone restrictions. *Studies in Language* 25:423–462.
- Gordon MK (2005): Intonational phonology of Chickasaw; in Jun SA (ed): *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford, University Press, pp 301–331.
- Grabe E (1998): Pitch accent realisations in English and German. *J Phon* 26:129–143.
- Grabe E, Post B, Nolan F, Farrar K (2000): Pitch accent realization in four varieties of British English. *J Phon* 28:161–185.
- Granström B, House D (2005): Audiovisual representation of prosody in expressive speech communication. *Speech Commun* 46:473–484.
- Grice M (1995a): Leading tones and downstep in English. *Phonology* 12:183–233.
- Grice M (1995b): The Intonation of Interrogation in Palermo Italian: Implications for Intonational Theory. Tübingen, Niemeyer.
- Grice M, Baumann S, Benz Müller R (2005): German intonation in autosegmental-metrical phonology; in Jun SA (ed): *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford, University Press, pp 55–83.
- Gronnum N (1989): Stress group patterns, sentence accents and sentence intonation in Southern Jutland (Sønderborg and Tønder) – with a view to German. *ARIPUK* 23:1–85.
- Gussenhoven C (1999): Discreteness and gradience in intonational contrasts. *Lang Speech* 42:283–305.
- Gussenhoven C (2004): *The Phonology of Tone and Intonation*. Cambridge, University Press.
- Hanson HM (2009): Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J Acoust Soc Am* 125:425–441.
- Hanssen J, Peters J, Gussenhoven C (2007): Alignment of the second low target in Dutch falling-rising pitch contours. *Proceedings of the 8th Interspeech Conference*, Antwerp, pp 2641–2644.
- Harrington J (2010): *The Phonetic Analysis of Speech Corpora*. Oxford, Blackwell Publishers.
- Harrington J, Kleber F, Reubold U (2008): Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *J Acoust Soc Am* 123:2825–2835.
- House D (2004): Pitch and alignment in the perception of tone and intonation: pragmatic signals and biological codes. *Proceedings of International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*, Beijing, pp 93–96.
- Igarashi Y (2002): Tak nazyvajemaja nejtralizacija intonacii – fonologičeskoje opisanije russkoj intonacii. *Bulletin Jpn Assoc Russ Scholarsh* 34:15–21.
- Igarashi Y (2005): How Many Falling Intonation Patterns in Russian? Categories of F0 Alignment. *Between Stress and Tone. Abstracts of IIAS Conference*, Leiden, pp 32–34.
- Igarashi Y (2006): Intonation patterns in Russian interrogatives – phonetic analyses and phonological interpretations; in Kawaguchi Y, Fonagy I, Moriguchi T (eds): *Usage-Based Linguistic Informatics 3: Prosody and Syntax*. John Benjamins, Amsterdam, pp 175–196.
- D’Imperio M (2001): Focus and tonal structure in Neapolitan Italian. *Speech Commun* 33:339–356.
- Jun SA (2005): *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.
- Jun SA (2014): *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford University Press.
- Knight RA, Nolan F (2006): The effect of pitch span on intonational plateaux. *J Int Phon Assoc* 36:21–38.
- Kohler KJ (1987): Categorical pitch perception. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallinn, pp 331–333.
- Kohler KJ (1991): Terminal intonation patterns in single-accent utterances of German: phonetics, phonology, semantics; in Kohler KJ (ed): *Studies in German Intonation*. AIPUK, vol 25, pp 295–368.
- Kohler KJ (1995): *Einführung in die Phonetik des Deutschen*. Berlin, Erich Schmidt Verlag.
- Ladd DR (2004): Segmental anchoring of pitch movements: autosegmental phonology or speech production? in Quené H, van Heuven V (eds): *On Speech and Language: Essays for Sieb B. Nootboom*. Utrecht, LOT, pp 123–131.
- Ladd DR (1996/2008): *Intonational Phonology*. Cambridge, Cambridge University Press.
- Ladd DR, Morton R (1997): The perception of intonational emphasis: continuous or categorical? *J Phon* 25:313–342.
- Lieberman M, Pierrehumbert J (1984): Intonational invariance under changes in pitch range and length; in Aronoff M, Oehrle R (eds): *Language Sound Structure*. Cambridge, MIT Press, pp 157–233.
- Lindblom B (1990): Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle W, Marchal A (eds): *Speech Production and Speech Modeling*. Dordrecht, Kluwer, pp 403–439.
- Maeda K, Venditti J (1998): Phonetic investigation of boundary pitch movements in Japanese. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*. Sydney, Australia.
- Makarova V (2003): Intonation features in categorization of Russian sentence type; in Kosta P, Błaszczak J (eds): *Investigations into Formal Slavic Linguistics IV*, pp 83–95.
- Makarova V (2007): The effect of pitch peak alignment on sentence type identification in Russian. *Lang Speech* 50:385–422.

- Meltzoff AN, Kuhl PK, Movell-Jones J, Sejnowski TJ (2009): Foundations for a new science of learning. *Science* 325:284–288.
- Morev LN (1983): *Shanskii Yazyk*. Moskva: Izdatelstvo Nauka.
- Mooshammer C, Geng C (2008): Acoustic and articulatory manifestations of vowel reduction in German. *J Int Phon Assoc* 38:117–136.
- Niebuhr O (2007): *Perzeption und Kognitive Verarbeitung der Sprachmelodie: Theoretische Grundlagen und Empirische Untersuchungen*. Berlin, de Gruyter.
- Niebuhr O (2012): At the edge of intonation: the interplay of utterance-final f0 movements and voiceless fricative sounds. *Phonetica* 69:7–27.
- Odé C (1989): *Russian Intonation: A Perceptual Description*. Amsterdam, Rodopi.
- Odé C (2005): Neutralization or truncation? The perception of two Russian pitch accents on utterance final syllables. *Speech Commun* 47:71–79.
- Odé C (2008): Transcription of Russian intonation ToRI, an interactive research tool and learning module on the internet; in Houtzagers P, Kalsbeek J, Schaeken J (eds): *Studies in Slavic and General Linguistics*. Amsterdam/New York, Rodopi, pp 431–449.
- Ohala J, Ewan W (1973): Speed of pitch change. *J Acoust Soc Am* 53:345(A).
- Ohl CK, Pfitzinger HR (2009): Compression and truncation revisited. *Proceedings of the 10th Interspeech Brighton, UK*, pp 2451–2454.
- Pierrehumbert JB (1980): *The Phonology and Phonetics of English Intonation*. PhD-Dissertation am MIT, Massachusetts.
- Pierrehumbert JB, Steele SA (1987): How many rise-fall-rise contours? *Proceedings of the 11th International Congress of Phonetic Sciences, Tallin*, pp 49.1.1–49.1.4.
- Pierrehumbert JB, Steele SA (1989): Categories of tonal alignment in English. *Phonetica* 46:181–196.
- Prieto P (2005): Stability effects in tonal clash contexts in Catalan. *J Phon* 33:215–242.
- Prieto P, van Santen J, Hirschberg J (1995): Tonal alignment patterns in Spanish. *J Phon* 23:429–451.
- Prieto P, Ortega-Llebaria M (2009): Do complex pitch gestures induce syllable lengthening in Catalan and Spanish? in Vigário M, Frota S, Freitas MJ (eds): *Phonetics and Phonology: Interactions and Interrelations*. Amsterdam, John Benjamins, pp 51–70.
- Prieto P, Torreira F (2007): The segmental anchoring hypothesis revisited: syllable structure and speech rate effects on peak timing in Spanish. *J Phon* 35:473–500.
- Rathcke T (2006a): A perceptual study on Russian questions and statements; in Harrington JM, Mooshammer C, Kleber F (eds): *Arbeitsberichte: AIPUK 37*, pp 51–62.
- Rathcke T (2006b): Relevance of f0 peak shape and alignment for the perception of a functional contrast in Russian. *Proceedings of the 3rd Conference on Speech Prosody, Dresden, Germany*, pp 65–68.
- Rathcke T (2013): On the neutralizing status of truncation in intonation: a perception study of boundary tones in German and Russian. *J Phon* 41:172–185.
- Rathcke T, Harrington J (2010): The variability of early accent peaks in standard German; in Fougeron C, Kühnert B, D’Imperio M (eds): *Laboratory Phonology 10. Phonology and Phonetics*. De Gruyter, Berlin/New York, pp 533–556.
- Schepman A, Lickely R, Ladd DR (2006): Effects of vowel length and ‘right context’ on the alignment of Dutch nuclear accents. *J Phon* 34:1–28.
- Silverman K, Pierrehumbert JB (1990): The timing of prenuclear high accents in English. *Lab Phon* 1:72–106.
- Steele SA (1986): Nuclear accent peak location: effects of rate, vowel and number of following syllables. *J Acoust Soc Am* 80(suppl 1):S51.
- Sundberg J (1979): Maximum speed of pitch changes in singers and untrained subjects. *J Phon* 7:71–79.
- Svetozarova ND (1982): *Intonatsionnaja Sistema Russkogo Jazyka*. [Intonational System of Russian.] Leningrad, Leningrad University Press.
- Uhmann S (1991): *Fokushphonologie. Eine Analyse Deutscher Intonationskonturen im Rahmen der Nicht-Linearen Phonologie*. Tübingen, Niemeyer.
- Watkins L (1984): *A Grammar of Kiowa*. Lincoln, Nebraska, University of Nebraska Press.
- Wenk R (1975): Die intonation; in Gabka K, Wiede E (eds): *Einführung in das Studium der Russischen Sprache*. Bd. 1: *Phonetik und Phonologie*. Leipzig, VEB, pp 175–198.
- Xu Y (1998): Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55:179–203.
- Xu Y, Sun X (2002): Maximum speed of pitch change and how it may relate to speech. *J Acoust Soc Am* 111:1399–1414.
- Yip M (1989): Contour tones. *Phonology* 6:149–174.
- Yu A (2003): Contour Tone Induced Lengthening in Cantonese. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain*, pp 2381–2384.
- Yu A (2010): Tonal effects on perceived vowel duration; in Fougeron C, Kühnert B, D’Imperio M, Vallée N (eds): *Papers in Laboratory Phonology 10*. Berlin, M. de Gruyter, pp 151–168.
- Zhang J (2002): *The Effects of Duration and Sonority on Contour Tone Distribution – A Typological Survey and Formal Analysis*. Routledge, New York.
- Zhang J (2004): The role of contrast-specific and language-specific phonetics in contour tone distribution; in Hayes B, Kirchner R, Steriade D (eds): *Phonetically Based Phonology*. Cambridge, Cambridge University Press, pp 157–190.