Eliciting speech with sentence lists – a critical evaluation with special emphasis on segmental anchoring

Lea S. Kohtz, Oliver Niebuhr

Department of General Linguistics, ISFAS, Christian-Albrechts-Universität zu Kiel, Germany

leakohtz@aol.com, niebuhr@isfas.uni-kiel.de

Abstract

We show on the basis of German that prosodic patterns change in the course of a traditional sentence-list elicitation. Two frequent methods are analyzed: sentence-frame and syntax-frame elicitations. While only the sentences of the sentence-frame elicitation show an increase in speaking rate, both elicitation methods cause a drastic reduction in the alignment variability of nuclear pitch-accent rises. So, the starting point for the idea of segmental anchoring, i.e. the characteristic stable alignment of L and H targets, could primarily be due to a training effect based on the continuous production of analogously constructed or identical carrier sentences. Detailed pitch-accent analyses also offer alternative interpretations for anchoring patterns. Methodologically, in order to avoid training effects in pitch-accent production, our findings suggest using the syntax-frame method and short sentence lists of 40 items or less.

Index Terms: elicitation, intonation, F0, segmental anchoring.

1. Introduction

It is now 15 years ago that Arvaniti, Ladd, and Mennen (cf. [1]) made the influential observation that the edges of (prenuclear) pitch-accent rises in Greek show a remarkably invariant alignment relative to segmental landmarks. More specifically, unless the duration of the accented syllable and the resulting temporal interval between the segmental landmarks was extremely short, the F0 minimum of the rise onset always occurred right before the accented-syllable onset, and the F0 maximum of the rise offset was consistently realized closely after the onset of the first postaccentual vowel. This phenomenon, for which Arvaniti et al. coined the term 'segmental anchoring', has since then been replicated and/or refined in numerous studies across many languages like English [2], Dutch [3], German [4,20], Russian [5], Japanese [6], and Mandarin Chinese [7].

There are to main reasons why the notion of segmental anchoring was so appealing to the intonational community. Firstly, it provides a very simple and powerful account of the phonetic variation of pitch accents and their phonological differences within and across languages. Secondly, and more importantly, segmental anchoring lends itself to an autosegmental-metrical interpretation, in which rise onset and offset represent separate tonal events – L and H – whose consistent alignments reflect a (secondary) phonological association of each tone with the segmental string, in addition to the (primary) phonological association of the entire pitch accent with the accented syllable. Thus, segmental anchoring suggests that it is the tonal alignment of L and H that speakers control across different segmental and prosodic contexts, whereas the F0 shape or slope in between L and H can vary largely across contexts, unlike predicted by contour models like the IPO model (cf. [8]). In this way, segmental anchoring was often considered to have resolved the level vs. contour debate in intonational phonology in favour of the level-based approach.

Apart from the fact that other contour models such as the Kiel Intonation Model, KIM [9], do not make strict claims about the invariance of shapes or slopes, but rather define contours and phonological differences between them in terms of fundamental curvature characteristics (e.g., peak vs. valley, or convex vs. concave) and the interplay of F0 elements (e.g., a change of element A causes a predictable change of elements B and C), segmental anchoring has also been subject of critical discuss-ion, cf. [10,11,12]. The present paper deals with an under-studied aspect of this discussion: elicitation methods, and how they can affect – and especially reduce – the alignment variability of the L and H targets of pitch-accent rises.

In a strict, literal understanding of the term, segmental anchoring presupposes that the alignments of L and H show a petty inter- and intra-subject variability; and in fact, for a matter-of-fact elicitation condition in which everything else may be considered equal, the observed variability is typically extremely low, particularly for L targets. It is understandable against this background that some papers on segmental anchoring report only grand means and refrain from providing any measures of dispersion (e.g., [4]). The matter-of-fact condition in which segmental anchoring is virtually exceptionally investigated is a list of syntactically analogously constructed isolated sentences. Speakers produce these sentences one after the other with a constant type of (broad-focus) pitch accent on grammatically determined, sonorous target words. Moreover, the sentences of a list are frequently repeated several times in separate blocks or in an overall randomized order; the lists, in turn, are additionally repeated at different speaking rates.

This paper is an initial attempt to critically evaluate this elicitation method. Based on our own experience with the elicitation of speech data, we wonder whether reading isolated sentences is really merely an efficient means to peel off or control all linguistic, paralinguistic, and micro-prosodic influences on pitch-accent alignment so that the underlying segmental anchoring can become observable. Isn't it also possible that the characteristic invariant L and H alignment of segmental anchoring is actually created or at least considerably facilitated by the sentence-list elicitation method? Speech production is essentially a muscular task; and as for every muscular task, repetitive training in a controlled, undisturbed environment reduces variability and increases precision, efficiency, and speed [21]. A list of 100-200 or even more isolated sonorous sentences looks like an ideal training ground. It allows speakers to (unconsciously) train their pitch accent production and timing, undisturbed by syntactic and prosodic variation, communicative purposes, and interference of voiceless segments, so that they can achieve a level of precision and constancy that would hardly be achievable in everyday speech.

Initial evidence in favour of this assumption comes from the study of [13] on the segmental anchoring of late LH* pitch accents in French. The authors of this study elicited a small corpus of less than 50 isolated sentences, as well as a paragraph corpus with a similarly small number of target sentences that were framed by syntactically and phonologically diverse

10.21437/Interspeech.2013-156

context sentences. Although the target sentences in this procedure are equally carefully controlled as in all other studies on segmental anchoring, it is obvious that the limited size of the two corpora in combination with the context sentences in the paragraph corpus does not allow speakers to intensively train their pitch-accent productions. On this basis, the pitch-accent alignment in [13:110] showed "a fair amount of variability […] within and across speakers […], in contrast to the very stable 'segmental anchors' found for other languages". This led to the assumption that tonal alignment in French is guided by wider anchorage areas in the segmental string rather than by specific segmental anchor points. Moreover, it is noted in [13] that "comparisons between the two corpora also reveal intra-speaker variability […]. There were almost no significant results for a given speaker that held across the two corpora".

We will present evidence in the following that the degree of sentence-list training can indeed make the difference between anchoring and anchorage: German native speakers read 200 isolated sentences in two frequently used list conditions, a sentence-frame and a syntax-frame condition. We measured for blocks of 20 sentences, how the general sentence prosody and the alignment details of nuclear pitch-accent rises change across the reading tasks. We found a number of changes, including a successive reduction of the alignment variability of L and H by up to 85% from the initial to the final 20 sentences of a list.

2. Method

Creating and recording our sentence lists was done with reference to 3 independent variables: Elicitation condition (ELI), Target-Word Type (QUANT), and list-wise repetition (REP).

Elicitation condition (ELI) means that we prepared two lists of sentences, which differed in how the target words were framed. In the first list, i.e. in the sentence-frame condition, the sentences were all completely identical except for the sentence-final target word. The target word was realized after the classic syntactically declarative context precursor "Das nächste Wort ist ___" (The next word is ___). In the second list of sentences, it was only the syntactic frame that remained constant across all sentences, whereas the words embedded in this frame varied. This construction principle, which we refer to as syntax-frame condition, is more similar those used by Arvaniti, Ladd, Xu, and others in previous studies of segmental anchoring. Our constant syntactic frame was syntactically declarative and constituted by a sequence of noun phrase, verb phrase (a single verb in 3rd pers. sg., present tense) and prepositional phrase. The latter phrase consisted of a preposition, an article, and – as in the first list – a sentence-final target word. Example sentences are "Die Woche beginnt am Montag" (The week starts on Monday) or "Die Mutter sitzt im Wagen" (The mother sits/waits in the car).

There were 20 target words and hence 20 sentences in each of the lists of the sentence-frame and syntax-frame conditions. All target words were similarly frequent German nouns. Moreover, all target words were disyllabic, with the initial CV(C) syllable bearing the lexical stress. When produced in the sentence contexts, the lexically stressed syllables were additionally associated with the nuclear pitch-accent rises whose alignment and segmental anchoring characteristics were analyzed. The pitch accents passed over into a steep terminal fall until the end of the utterance. Based on the variable QUANT, the target words fell into two groups. Ten target words had a short vowel in the stressed/accented syllable, followed by a voice-less obstruent in the syllable coda. The syllable onset was al-ways a voiced consonant. The other ten target words had a long vowel or a diphthong in the stressed/accented syllable. The long vowels or diphthongs were also preceded by a voiced consonant. If there was a coda consonant, it was also voiced.

The recruited speakers were 4 male and 5 female native speakers of German. They were between 23 and 39 years old and lived their whole life in Northern Germany. Most of them were good friends of the first author and students of Empirical Linguistics at the University of Kiel.

The recordings were made digitally (88.2 kHz, 24-bit) in separate sessions in the sound-treated Speech Lab of the Department of General Linguistics. At the be beginning of each session, the speaker was seated in front of a computer monitor and instructed to read the target sentences, which would be displayed one by one on the screen, in the form of separate isolated utterances at normal (i.e. convenient) speaking rate and loudness levels. They were also told that there would be different types of sentences and that each sentence would occur several times during a session.

After all 20 sentences of a list had been realized, the list was presented again with a different randomization for a total of 10 times. The 10 differently randomized repetitions constituted the REP variable. A small break was inserted after each repetition. After the 10th repetition, a longer break was inserted, before the 20 sentences of the other list were read in 10 differently randomized repetitions. About half of the speakers started with the list of the sentence-frame condition. The other half started with the list of the syntax-frame condition.

Before the acoustic analyses were started, all recordings were resampled to 22 kHz and 16-bit. The measurements were taken in PRAAT (www.praat.org) using the default analysis settings. Two types of measurements were taken. The first type concerned general characteristics of sentence prosody. These included the overall duration and mean intensity of the sentences, as well as the intonational variability in terms of the standard deviations for the L and H alignments of the nuclear pitch-accent rises. The alignments of L and H were measured relative to those segmental landmarks whose relevance had been attested in many previous studies, cf. [14,15,16]. In the case of L, this is the accented-syllable onset. In the case of H, the accented-vowel onset served as segmental landmark. It must be noted that the general sentence characteristics were analyzed repetition wise. That is, the values of the 20 sentences of each repetition were integrated in means or standard deviations, which then served as input for further statistical analyses. Over and above the general characteristics of sentence prosody, the acoustic analysis also included some timing details of the nuclear pitch accent rises, such as the duration of the pitch-accent rise, the distance of L relative to the syllable onset, and the distances of H relative to the two boundaries of the accented vowel.

3. Results

3.1. General characteristics of sentence prosody

A three-way MANOVA was used to test the effects of the fixed factors Repetition (REP, 10 levels), Elicitation Condition (ELI, 2 levels) and Target-Word Type (QUANT, 2 levels) on the acoustic parameters sentence intensity, sentence duration, and intonational variability. The MANOVA yielded significant main effects of all three factors (REP: F[36,1280]=13.2, p<0.001, η_p^2=0.27; ELI: F[4,317]=7.9, p<0.001, η_p^2=0.09; QUANT: F[4,317]=480.4, p<0.001, η_p^2=0.86). The three-way interaction REP x ELI x QUANT was not significant.

The significant main effects primarily stem from sentence duration and intonational variability. The only significant difference in sentence intensity concerned the Target-Word Type and was due to the fact that sentences with long-vowel target words had a slightly (about 1 dB) higher intensity than sentences with short-vowel target words (F[1,320]=6.5, p<0.012, η_p^2=0.02). Sentence intensity varied between 50-60 dB.

Sentence durations were overall longer in the syntax-frame than in the sentence-frame condition (F[1,320]=29.3, p<0.001, η_p^2=0.08). A similar difference was found for the comparison of sentences with long-vowel and short-vowel target words (F[1,320]=16.5, p<0.001, η_p^2=0.05). Apart from these expectable findings that reflect intrinsic duration differences due to syntactic complexity or vowel quantity, sentence duration also decreases significantly the more often the blocks of 20 sentences were repeated (F[9,320]=5.9,p<0.001, η_p^2=0.14). However, this does not apply to the syntax-frame condition, whereas in the sentence-frame conditions, the sentences of repetition 10 are on average about 20% (or 500 ms) shorter than the sentences of repetition 1 for both target-word types. This differential effect of Elicitation Condition on sentence duration is reflected in the MANOVA in a significant interaction between Elicitation Condition and Repetition (F[36,1280]= 1.6, p<0.014, η_p^2=0.04).

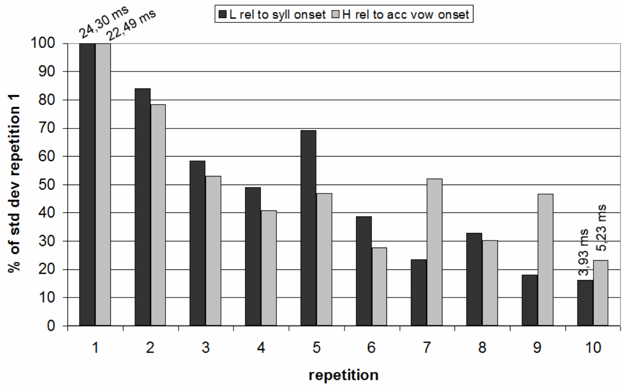

Figure 1: *Standard deviations for the alignment of the nuclear L and H tones in the accented syllable across repetitions 1-10, shown for the short-vowel target words of the syntax-frame condition in terms of percentages relative to repetition 1; each bar n=90.*

As regards intonational variability, the standard deviations for the alignment of the nuclear L and H pitch-accent tones both differ highly significantly between the two Target-Word Types (L: F[1,320]=1021.3, p<0.001, η_p^2=0.76; H: F[1,320]= 870.9, p<0.001, η_p^2=0.73). The alignment variability of L and H was higher for accented syllables with long-vowel nuclei. Addi-tionally, and more importantly, how much the alignment of L and H varied was also greatly affected by Repetition (L: F[9, 320]=204.2, p<0.001, η_p^2=0.85; H: F[9,320]=105.0, p<0.001, η_p^2=0.75). More specifically, as is exemplified in Figure 1 for the short-vowel target words of the syntax-frame condition, the standard deviations of both L and H decrease considerably across repetitions 1-10 from about 23-24 ms to only about 4-5 ms. So, nuclear L and H tones in the sentences of repetition 10 show 75-85 % less alignment variation relative to their respective syllable- or vowel-onsets than in the sentences of repetition 1. This stabilization effect occurred independently of Elicitation Condition, but was slightly weaker for the long-vowel than for the short-vowel target words, which resulted in a significant interaction between Repetition and Target-Word Type (F[36,1280]=2.6, p<0.001, η_p^2=0.07).

3.2. Details of nuclear pitch-accent alignment

In order to get an impression of the details of nuclear pitch-accent alignment, we looked at the two corner points of each elicitation condition, i.e. the 20 sentences of repetitions 1 and 10. T tests (*df*=89) with Bonferroni-corrected p values were used to compare the L and H alignments of the nuclear pitch accents between repetitions 1 and 10. The results are summarized in Table 1 and schematically illustrated in Figure 2.

Table 1: *Mean durations of the nuclear pitch-accent rises (LH dur) and mean alignments of L relative to the syllable onset (minSyll) and of H relative to the vowel boundaries (maxVon, max Voff) in repetitions 1 and 10; t-test statistics for comparisons of repetitions 1 and 10, displayed separately for elicitation condition and short-vowel or long-vowel target words.*

ELI	QUANT	nucl. ris	REP.1	REP.10	t	p
sentence frame	SV	LH dur	104.9	92.9	14.9	<0.001
	SV	minSyll	-12.1	-9.3	1.4	n.s.
	SV	maxVon	39.5	38.7	-1.1	n.s.
	SV	maxVoff	42.6	28.3	11.4	<0.001
	LV	LH dur	145.8	127.1	5.8	<0.001
	LV	minSyll	-2.3	-3.7	0.6	n.s.
	LV	maxVon	91.8	86.0	-0.4	n.s.
	LV	maxVoff	57.2	34.8	10.9	<0.001
syntax frame	SV	LH dur	89.6	75.4	11.6	<0.001
	SV	minSyll	3.4	5.5	0.9	n.s.
	SV	maxVon	26.7	15.8	14.8	<0.001
	SV	maxVoff	29.3	41.6	-9.6	<0.001
	LV	LH dur	96.3	80.6	15.0	<0.001
	LV	minSyll	-4.0	-6.9	-1.5	n.s.
	LV	maxVon	58.2	38.7	8.2	<0.001
	LV	maxVoff	65.8	76.8	-3.43	<0.01

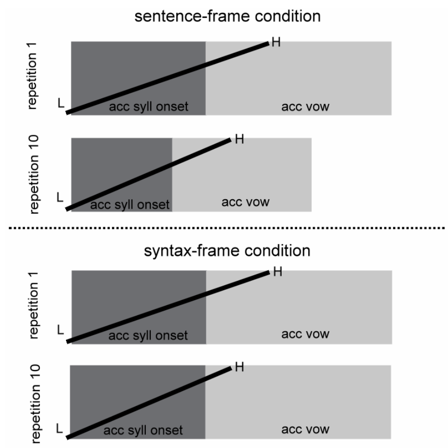

Figure 2: *Schematic illustration of the changes in the nuclear pitch-accent rises between repetitions 1 and 10 of the sentence-frame (top panel) and the syntax-frame (bottom panel) condition, based on Table 1.*

The nuclear pitch-accent rises in the sentence-frame condition, start in both repetitions immediately before the accented-syllable onset. However, the rises in repetition 10 were shorter than those in repetition 1. Taking into account that the sentences in the sentence-frame condition are produced about 20%

faster in repetition 10 than in repetition 1, it becomes understandable why the nuclear H remained stably aligned about 40 ms after the accented-vowel onset (cf. Tab.1), despite the difference in rise duration, whereas the distance between H and the accented-vowel offset decreases at the same time significantly from repetition 1 to 10 (cf. Fig.2, top panels).

The intonational pattern is overall the same in the syntax-frame condition. But, since sentence durations in this elicitation condition do not decrease in the same way, the shorter rise durations in repetition 10 shift the H tone significantly closer to the accented-vowel onset and further away from the accented-vowel offset (cf. Fig.2, bottom panels).

4. Discussion

It is common scientific practice that not only the results of a study become subject of critical discussion, but also the method with which they were obtained. The key questions are those of external and ecological validity; for example, to what extent do research methods affect or change the studied objects, and do the analyzed objects (still) correspond to the intended objects? These questions have of course always played a role in phonetic research, and the last thing we want to imply is that phonetic researchers do not carefully select and apply their elicitation methods. On the other hand, it is probably not exaggerated to say that elicitation methods are primarily selected and applied on intuitive grounds. We have little, if any, empirical data that could tell us, how different elicitation methods or conditions (linguistic diversity and orthographic representation of reading material, amount of multi-modal contextualization, read or unscripted speech, table or head-mounted micrphones, monologues or dialogues, physical presence or absence of a dialogue partner, etc.) affect the segmental and prosodic patterns that speakers produce, which method is to be preferred for which question, and how far we can push and vary our methods before we create artifacts. Against this background, our study is an attempt to initiate a separate methodology-oriented line of phonetic research. An integral part of such a line of research is that speakers are not regarded as "vending machines" that generate the desired signal by paying and pressing a button. This technical perspective on elicitation methods is useful, but must be complemented by social and meaning-related aspects of speech communication.

We started our methodology-oriented line of phonetic research from the popular concept of segmental anchoring and compared two frequent types of sentence-list elicitation methods by means of German speakers. Firstly, we found in terms of general prosodic characteristics that the sentence-frame method (based on the carrier sentence "The next word is ___") causes a considerable decrease in sentence duration, which is due to an increase in speaking rate. The final 20 sentences were uttered about 20 % faster than the initial 20 sentences of the list. This elicitation effect was not found for the syntax-frame method, probably due to the greater diversity of this reading task, in which the carrier sentences were just analogously constructed, but not identical. Secondly, we found for both elicitation methods a drastic reduction in the variability of the alignment of the nuclear pitch-accent rises. For the short-vowel target words, rise onset (L) and rise offset (H) in the final 20 sentences were up to 85 % more stably aligned relative to the onsets of the accented-syllable (L) or the accented-vowel (H) than in the initial 20 sentences. For the long-vowel target words, the variability reduction across the reading tasks still amounted to about 50 %.

It is true that most studies on segmental anchoring actually analyzed prenuclear pitch accents instead of the nuclear accents that we analyzed here. However, we think that this point is irrelevant. The stabilizing effect that we found for L and H across the ten repetitions is not predicted by segmental anchoring, irrespective of where in the intonation phrase the pitch accent is located. If segmental anchoring was a basic force in pitch-accent production, reflecting phonological structure, then there should be a high precision and invariance in L and H alignment from the first sentence onwards.

So, in addition to recent critical discussions about segmental anchoring in [10,11,12], our finding suggests that segmental anchoring is primarily a training effect, created or at least considerably facilitated by the continuous elicitation of isolated and phonetically highly controlled sentences, which reduce speech production from a means of social interaction and information transfer to a monotonous muscular exercise. The alignment details of the analyzed nuclear pitch accents point in the same direction. In the sentence-frame condition, the sentences with a higher speaking rate show shorter pitch-accent rises so that L and H can maintain their relative distances to the syllable or vowel onsets. However, the pitch-accent rises are also shorter in the syntax-frame condition, despite the constant speaking rate in this condition. Therefore, it seems doubtful that the shorter rise duration in the sentence-frame condition reflects segmental anchoring. Rather, is it reasonable to assume that the shorter rises in both conditions are merely a by-product of the sentence-list training, which increases not just the precision, but also the speed of muscular movements, including those that underlie the production of F0. On the whole, the alignment details of L and H agree well with previous findings on Northern Standard German [15,18,20].

In summary, returning to the terminology of [13]: sentence-list elicitations can turn an anchorage area into an anchor point for intonational targets. Note that this conclusion stresses at the same time that timing phenomena like the temporal coordination among pitch-accent targets and systematic alignment differences between pitch-accent categories do of course exist, and that this more general understanding of segmental anchoring is in no way undermined by our findings. The main stabilization of the alignment of L and H occurred after the second repetition, i.e. after the 40th sentence. This agrees well with the fact that studies that analyzed less or only slightly more than 40 read-speech sentences still found a high degree of pitch-accent variability [13], including a trade-off between alignment and shape properties which can easily mask grand means [17]. Since also the increase in speaking rate in the sentence-frame condition set in after the 40th sentence, we think that lists of isolated sentences are still a suitable elicitation method as long as the lists are kept short enough. Whether measures like contextualization or a physically present addressee allow longer lists must be tested in follow-up studies.

Furthermore, our findings suggest that sentence-frame elicitation may be a proper means to analyze target words at different speaking rates. Eliciting a fast rate by asking speakers to produce utterances "as fast as they can" is a questionable strategy, since it is likely to induce mental stress, which then interferes with the analyzed speech production patterns. In fact, a raised F0 level, which is frequently noted in the analysis of deliberately fast utterances, is a typical feature of speech under mental stress [19]. This is another example, how necessary and worth-while it will be to learn more in future studies about the advantages and limitations of the elicitation tasks that we use.

5. References

[1] Arvaniti, A., Ladd, D.R., Mennen, I.," Stability of tonal alignment: Greek prenuclear accents", J. Phon. 26, 3-25, 1998.

[2] Ladd, D. R., Faulkner, D., Faulkner, H., Schepman, A., "Constant 'segmental' anchoring of f0 movements under changes in speech rate", JASA 106, 1543-1554, 1999.

[3] Ladd, D.R., Mennen, I., Schepman, A, "Phonological conditioning of peak alignment in rising pitch accents in Dutch", JASA 107, 2685-2696, 2000.

[4] Atterer, M., Ladd, D.R., "On the phonetics and phonology of segmental anchoring of f0: evidence from German", J. Phon. 32, 177-197, 2004.

[5] Igarashi, Y., "'Segmental anchoring' of f0 under changes in speech rate: Evidence from Russian", Proc. 2nd Speech Prosody, Nara, Japan, 25-28, 2004.

[6] Ishihara, T., "A phonological effect on tonal alignment in Tokyo Japanese", Proc. 15th ICPhS, Barcelona, Spain, 615-618, 2003.

[7] Xu, Y., "Consistency of tone-syllable alignment across different syllable structures and speaking rates", Phonetica 55, 179-203, 1998.

[8] Adriaens, L.M.H., "Ein Modell deutscher Intonation", PhD thesis, Technical University Eindhoven, The Nethelands, 1991.

[9] Kohler, K.J., "Paradigms in experimental prosodic analysis: From measurements to function", Sudhoff et al. [Eds], Methods in empirical prosody research, 123-152, Berlin: deGruyter, 2006.

[10] Prieto, P., Torreira, F., "The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish", J. Phon. 35, 473-500, 2007.

[11] Cho, H., "The timing of phrase-initial tones in Seoul Korean: a weighted-constraint model", Phonology 28, 293-330, 2011.

[12] Ladd, D.R., "Segmental anchoring of pitch movements: autosegmental phonology or speech production?" In Quené, van Heuven [Eds], Speech and Language: Studies for Sieb G. Noteboom, 123-131, Utrecht: LOT, 2004.

[13] Welby, P., Loevenbruck, H., "Anchored down in Anchorage: Syllable structure and segmental anchoring in French", Italian J. Ling. 18, 73-124, 2006.

[14] Niebuhr, O., Kohler, K.J., "Perception and Cognitive Processing of Tonal Alignment in German", Proc. International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages, Beijing, China, 155-158, 2004.

[15] Gartenberg, R., Panzlaff-Reuter, C., "Production and perception of f0 peak patterns in German", AIPUK 25, 29-115, 1991.

[16] Niebuhr, O., "The role of the accented-vowel onset in the perception of German early and medial peaks", Proc. 3rd International Conference of Speech Prosody, Dresden, Germany, 109-112, 2006.

[17] Niebuhr, O., M. D'Imperio, B. Gili Fivela, F. Cangemi, "Are there 'shapers' and 'aligners'? Individual differences in signalling pitch accent category", Proc. 17th ICPhS, Hong Kong, China, 120-123, 2011.

[18] Niebuhr, O., Ambrazaitis, G.I., "Alignment of medial and late peaks in German spontaneous speech", Proc. 3rd International Conference of Speech Prosody, Dresden, Germany, 161-164, 2006.

[19] Johannes, B., P. Wittels, R. Enne, G. Eisinger, C.A. Castro, J.L. Thomas, A.B. Adler, R. Gerzer, "Non-linear function model of voice pitch dependency on physical and mental load", Eur J Appl Physiol 101, 267–276, 2007.

[20] Kleber, F., Rathcke, T., "More on the "segmental anchoring" of prenuclear rises: Evidence from East Middle German", Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 583-586, 2008.

[21] Schmidt, R., Lee, T.D., "Motor control and learning – A behavioral emphasis", Champaign, Human Kinetics, 2011.