Semantic-context effects on lexical stress and syllable prominence

Felicitas Kleber[1] *& Oliver Niebuhr*[2]

[1]Institute of Phonetics and Speech Processing, University of Munich, Germany
[2] Department of General & Comparative Linguistics, University of Kiel, Germany
kleber AT phonetik.uni-muenchen.de; niebuhr AT linguistik.uni-kiel.de

Abstract

In the present study we investigated the effect of semantic context on the perception of words that differ only in the lexical stress pattern. Our study was based on the disyllabic German word *AUGUST*. Depending on whether the lexical stress is on the first or the second syllable the word refers to either a name or a month. By means of a forced-choice identification experiment we tested to what extent the lexical stress-position is triggered by the semantic context and by local phonetic cues. The stimuli that constituted a 7-step continuum from ***'Au****gust* to *Au****'gust*** were appended to different word lists containing another name and another month, i.e. context words that were semantically related to either the name ***'Au****gust* or the month *Au****'gust***. Results showed that the perceptual boundary between ***'Au****gust* and *Au****'gust*** was shifted according to the semantics of the adjacent context word. This semantic-context effect was present for both ambiguous and clear phonetic prominence cues in *AUGUST*.

Index terms: speech perception, prominence, stress, German.

1. Introduction

Speech perception is not a local identification and successive concatenation of meaning-related units in order to build up the message conveyed by the speaker. This fact is known for a long time. And although the full range of its implications is only revealed step by step, the steps that have already been made underline the importance of *phonetic* and *signal-external* communicative contexts, in this way also challenging traditional dichotomies like 'segmental' vs. 'suprasegmental' or 'categorical' vs. 'gradual'.

As regards the phonetic context, for example, phonemic units do not only manifest themselves segmentally, i.e. in a bundle of local features. Rather, phonemic features like palatality or nasality can spread widely over the surrounding segmental context, cf. [1,2]; and these suprasegmental exponents can persist or are even enhanced when the corresponding speech segments themselves disappear due to assimilation or elision, cf. [3,4]. In the opposite direction the representations of intonational units like pitch accents and boundary tones go beyond local fundamental-frequency (F0) changes and involve the surrounding segmental context. For example, truncated rising or falling boundary tones are continued in the pitch created by variation in the spectral-energy distribution of utterance-final fricatives, cf. [5]. Since the hearer is able to use such mutual variations in the coding of phonemic and prosodic units (cf. [6,7,8,9,10]), speech communication can be successful even when the phonetic signal in the segmental or the suprasegmental domain is ambiguous or deviates considerably from the (claimed) phonological basis.

Furthermore, in addition to the phonetic context, hearers integrate signal-external context when perceiving speech. This holds in the same way for words and pitch accents, cf. [11,12]. For instance, [13] demonstrated that hearers restore sounds that were masked by noise when they identify words like *"__eel"*. However, if they perceive *"wheel"* or *"peel"* depends on the semantic context of the utterance. Utterances that dealt with oranges triggered *"peel"* identifications, whereas hearers heard *"wheel"* when the utterance was about an axle. Similarly, spoken-word identification is influenced by the semantics of visual stimuli like text or pictures, cf. [14,15].

In summary, speech-perception research creates a picture of speech communication in which the hearer is not a machine that passively soaks up and decodes the incoming (acoustic-) phonetic signal. Rather, speech communication is always done over time in a particular (and mostly noisy) environment and with changing shared knowledge between the interlocutors. Within this situational framework, the hearer is a creative interpreter who makes use of all cross-modally available, contextual information to decode the current phonetic input. The present study will contribute to flesh out this picture further with reference to one of the most context-dependent and central phenomena in speech: *Prominence*.

Prominence refers to the extent to which a syllable is perceived as standing out against the surrounding ones. Patterns of higher and lower prominences, are, among others, involved in the signalling of focus, pitch accent, emphasis, phrasing, speech rhythm, as well as lexical-stress position, cf. [17]. Phonetically, prominence is based on local changes (typically, but not necessarily increases) of parameters like f0, (segment) duration, and intensity, cf. [18,19]. But other signal-external factors such as the preceding and even the following rhythmic and intonational context ([20,21]), as well as the hearer's knowledge about the grammatical categories of words and the operation of the speech-production apparatus can considerably influence prominence perceptions and completely override local phonetic cues, cf. [12,22,23,24].

Although prominence was intensively studied across languages, it has not been shown so far that the perception of prominence also includes a semantic-context effect, analogous to the *"wheel"-"peel"* effect in the perception of words. However, this is an important issue, since the coding of many linguistic and paralinguistic meanings involve prominence (cf. above), and since the signal-external factor 'semantic context' is constantly present in speech communication. Therefore, our perception study aims at determining (a) *if* and (b) *to what extent* prominence is affected by the semantic context. For this purpose, we exploit the fact that – analogous to English – German has minimal pairs of words that are just differentiated by the lexical-stress position. This position is perceptually linked with a higher syllable prominence. That is, lexical stress triggers higher prominence, and prominence cues, in turn, cue lexical-stress positions. Our experiment was based on the disyllabic target-word pair ***'Au****gust* vs. *Au****'gust***, [ˈaʊ.gʊstʰ]. With lexical stress and hence higher prominence on the first syllable, the word represents a male *name*. If the second syllable bears the lexical stress and is hence more prominent, the word refers to the eighth *month* of the year. In order to address the main aim (a) that concerns the 'if', the target disyllable *AUGUST* was embedded into different semantic contexts. For determining the strength of the potential

semantic-context effect according to aim (b), we created a phonetic prominence-cue continuum in the target disyllable which shifts the perceived lexical-stress successively from the first to the second syllable via a number of ambiguous stimuli. Then, we looked at interactions of local prominence cues and semantic context on the perceived lexical-stress position. The lexical-stress position was measured in terms of word identification, i.e. name or month.

A similar approach was already used in a previous perception experiment in which we just found an effect of the local prominence cues, but not of the semantic context. However, this could have been an artefact of the method. In the previous perception experiment, the target disyllable *AUGUST* was presented constantly at initial position in the stimulus. The stimulus itself was preceded by utterances whose semantic contexts matched with either the name or the month in the target disyllable. However, due to the predictable occurrence of *AUGUST*, it was possible for the subjects in the course of the experiment to ignore the context and to narrow their focus down to the target disyllable and its local prominence or lexical-stress cues. It is obvious that semantic-context effects cannot show up under these circumstances. Accounting for this shortcoming of our previous experiment, the present experiment used word lists of different length to which the target disyllable were attached. This strategy makes it harder for subjects to predict the occurrence of the target disyllable. On this basis, the following hypotheses were put forward.

(H1) The semantic context has an effect on the lexical-stress and prominence patterns and hence on the word that is identified in the target disyllable. Specifically, a name/month context supports lexical-stress and higher prominence on the first/second syllable of AUGUST, respectively.

(H2) The semantic-context effect is stronger, if the local phonetic cues to prominence and lexical stress in the target disyllable are ambiguous.

The first hypothesis (H1) refers to the fact that [12,20,21] have already shown by means of variations in rhythm and intonation that prominences and lexical-stress positions can in principle be affected by the preceding signal-external context. Moreover, the *"wheel"* vs. *"peel"* example shows that semantic-context effects can in principle occur in speech perception. The second hypothesis (H2) takes into account that it is a frequent finding in (speech) perception that contextual factors become more influential if local cues are less clear (e.g., [10, 12,20,21,24]).

2. Method

Our hypotheses were tested within a forced-choice identification experiment.

2.1. Stimulus generation

We created a 7-step stimulus continuum in between ***'Au**gust* and *Au**'gust*** from an ambiguous *AUGUST* which was produced naturally with two equally prominent syllables (i.e. with stress clash) by a speaker of Standard German. The stimuli differed in the duration ratio of the first vowel (V_1, [aʊ]) to the second vowel (V_2, [ʊ]). For a clear percept of ***'Au**gust* in stimulus 1, V_1 was lengthened by about 50ms whereas V_2 was equally shortened. By contrast, in order to create a clear percept of *Au**'gust*** for stimulus 7, V_1 was shortened by about 50ms while V_2 was equally lengthened, cf. Figure 1. In the further intermediate duration ratios of stimuli 2-6, in which the complementary changes in V_1 and V_2 durations were done in smaller steps of about 10ms, the perceived lexical-stress in the isolated *AUGUST* items is shifted successively from the first syllable to the second one, via a number of perceptually ambiguous stimuli, Table 1.

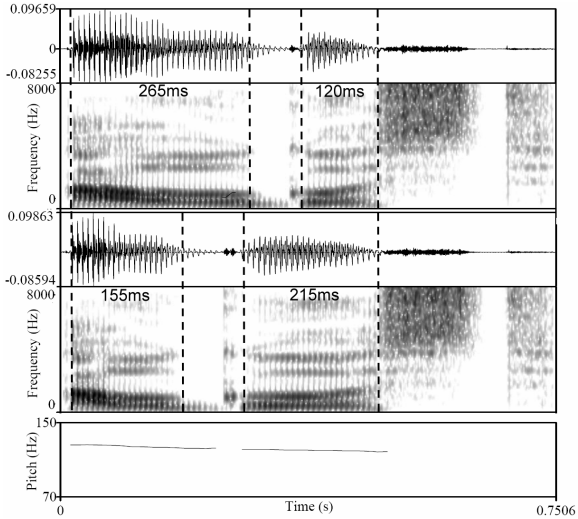

Figure 1: *Oscillograms and spectrograms showing the contrast in vowel duration between the extreme stimuli of the created continuum from **'Au**gust (top) to Au**'gust** (middle). The bottom panel displays the constant F0 decline of all stimuli.*

Table 1. *Vowel durations of the stimuli from the continuum.*

Stimulus (Semantics)	Vowel duration [ms]		duration ratio V_1/V_2 (in %)
	V_1	V_2	
1 (name)	265	120	69/31
2	237	144	62/38
3	224	156	59/41
4	210	168	56/44
5	197	180	52/48
6	183	192	49/51
7 (month)	155	215	42/58

The speaker also produced the words ***'Frie**drich* (a name) and ***'Ju**li* (July). *'**Frie**drich* and ***'Au**gust* are frequently combined first names in German and ***'Ju**li* is the month preceding *Au**'gust***. Together, the two months are associated with summer or summer holidays. Moreover, ***'Frie**drich* and ***'Ju**li* are both disyllables as *AUGUST*, but with constant lexical stress on the first syllable. Hence, context effects of lexical stress and the resulting rhythmic patterns are controlled, cf. [17,18]. The two words ***'Frie**drich* and ***'Ju**li* were combined in different orders to contextualize the target disyllable *AUGUST*. In the first list – the month list – the two words appeared in the order *Friedrich, Juli, Juli*. This context could trigger a bias towards *Au**'gust***, because the month context outnumbered the name context and was adjacent to the target disyllable. The second list – the name list – was compiled to trigger the name ***'Au**gust*. To this end, we put the words of the second list in the order *Juli, Friedrich, Friedrich*. We derived 4 additional lists from these two basic lists by adding the two context words in both possible orders at the beginning of the month and the name list. The 7 stimuli were then attached to each of the 6 different word lists, resulting in 42 different lists such as, for example, *Juli, Friedrich, Juli, Friedrich, Friedrich, AUGUST.*

It was necessary to use 3-item and 5-item context lists in order to make the position of the target disyllable in the list unpredictable. Moreover, our 42 experimental lists were complemented by a large number of *filler* lists. In some of them *'August* and *Au'gust* occurred in non-final position and/or after a *single* item of *Friedrich* or *Juli*. In this way, subjects had to

pay attention to both context words and target disyllables (i.e. the 7 stimuli).

The reason for keeping f0 at a constant (slightly declining) level in both the stimuli and the context words, although f0 is the stronger cue to lexical stress in German (cf. [25]), was that the word lists were compiled of single words produced in isolation. If each of these words had a different and/or variable f0 contour, this would have introduced phrasal and rhythmic structures into the concatenated context words and stimuli. Apart from the fact that these structures would have varied from list to list, which yields another experimental variable, it was crucial for the reliability and comparability of the context-ualization that the subjects were able to perceive the concatenated items as an entire list without any complex substructure.

2.2. Subjects, experiment, and data analysis

Eighteen Standard German speakers (9 female) participated in the experiment. The subjects' age ranged from 18 to 56 years.

The experiment was run on a computer using the e-prime software version 2.0, cf. [26]. Ten repetitions of the 14 stimulus lists with three context words and five repetitions of the 28 stimulus lists with five context words were presented in random order over headphones together with the filler lists. The subjects were asked to judge whether they perceived the last word in each list as the name ***'Au****gust* or the month *Au**'gust***. The experiment lasted 2 hours, but the subjects were free to take breaks after every 84 stimulus lists. It was up them how long their break lasted.

First, we carried out a repeated measures ANOVA with Context (within-subject factor with 2 levels: month list vs. name list) and V_1/V_2 Ratio (within-subject factor with 7 levels: stimuli 1-7) as independent variables. The 'month'-Responses served as the dependent variable. Secondly, we calculated the category boundary (i.e. the point in the continuum for which 'month' responses were 50%) for each subject using probit analysis and a paired t-test was performed to test whether Category Boundaries obtained for the two lists differed.

3. Results

The results of three subjects, who responded to the entire continuum with 'month', were excluded from the analysis. The responses of the remaining 15 subjects to the different lists were grouped into the two basic list types, Month and Name, after it was ensured that the responses did not differ with respect to the length of the list.

Figure 2 summarizes the responses as a function of decreasing V_1/V_2 duration ratios across the *AUGUST* continuum. There is a clear shift in the perceived lexical-stress from V_1 to V_2, i.e. stimulus 1 was perceived as the name ***'Au****gust* (86% in both contexts) and stimulus 7 was clearly perceived as the month *Au**'gust*** (90% vs. 87% in the month and name context, resp.). So, in line with [25], the V_1/V_2 duration ratio proofed to be a strong acoustic cue to lexical stress in German. The corresponding ANOVA revealed a significant effect of V_1/V_2 Ratio ($F(6,84) = 94.041$, $p < 0.001$), as well as a significant effect of Context ($F(1,14) = 5.5263$, $p < 0.05$), but no significant interaction for Context*V_1/V_2 Ratio. This means that responses to all stimuli - independently whether the V_1/V_2 ratio cue is ambiguous or clear - are affected by the semantic context. We derived logistic regression curves from the psychometric functions in order to determine the category boundaries on the acoustic continuum. The mean category boundaries for the month (=3.7) and the name context (=4.1) are superimposed on the regression curves in Figure 3. A paired t-test revealed no significant difference ($t = -1.4812$, $df = 14$, $p = 0.1607$) between the category boundaries in the name vs. the month context.

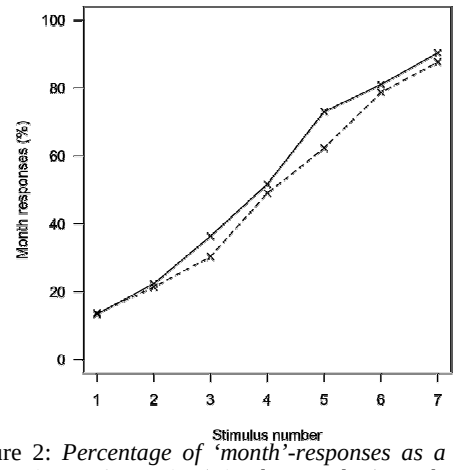

Figure 2: *Percentage of 'month'-responses as a function of decreasing V_1/V_2 ratio (stimulus number) to the month list (solid) and the name list (dotted).*

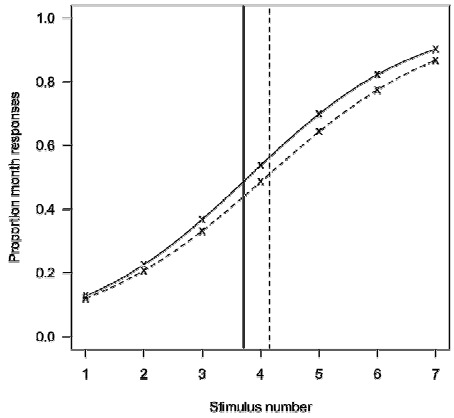

Figure 3: *Regression curves derived from the identification functions of 'month'- responses to the month list (solid) and the name list (dotted). The vertical lines show the mean category boundaries in the **'Au**gust-Au**'gust** continuum following the month list (solid) and the name list (dotted).*

4. Discussion and Conclusion

Whether the listeners identified the target disyllables as either the initially-stressed ***'Au****gust* or the finally-stressed *Au**'gust*** was not only a matter of the duration ratio that was used as the local phonetic cue to syllable prominence and hence to lexical-stress position. The perception of ***'Au****gust* and *Au**'gust*** and hence of the lexical-stress position was also determined by the semantics of the preceding context word, which was either the name ***'Fried****rich* that is frequently combined with the name ***'Au****gust* or the month ***'Ju****li* that is adjacent to the month *Au**'gust***. The context word significantly supported the perception of the semantically-related target disyllable. So, compared with the name ***'Fried****rich* the month ***'Ju****li* increased *Au**'gust*** identifications. This effect of the signal-external semantic context on lexical stress and syllable prominence is in line with hypothesis H1. However, as against the second hypothesis H2, we found no evidence for an interaction between the phonetic and the semantic-context factors. The effect of the preceding context word on lexical stress and syllable prominence was not larger in the middle of the stimulus continuum, where the duration ratios were more ambiguous prominence or stress cues than at the ends of the stimulus continuum.

What are the implications of these findings for speech communication? Although our experimental task abstracts quite a bit away actual speech communication, the listeners in our experiment decoded strings of words that were to a certain extent unpredictable, and that had internal meaning relationships. These are also fundamental characteristics of the speech code. Meaning relationships exist in enumerations (which are close to our stimuli), between verbs or adjectives and nouns (e.g., certain verbs/adjectives rule out certain nouns), and they are even contained in syntactic congruency. Thus, by showing that these fundamental meaning relationships are not only exploited to restore or to disambiguate words or pitch accents, but that they are simultaneously used as cues to lexical-stress and hence prominence perception, our study contributes another building block to the picture that shows the listener as an active, knowledge- and expectation-driven interpreter rather than a passive decoder of the acoustic phonetic signal, cf. also [11,12,13,14,15,16]. Moreover, our findings further support the view of prominence as a perceptual phenomenon with a considerable cognitive (i.e. top-down) momentum. Finally, with regard to the relevance of our findings for speech communication, it is worth noting that our semantic-context effect occurred also for disyllables with *clear* phonetic prominence or lexical-stress cues, as they may be expected in real speech production. It is possible that this robustness against local phonetic cues is a characteristic of meaning-based context effects in speech perception. For example, [21] found effects on syllable prominence and lexical stress for rhythmic and intonational contexts. However, while the rhythmic-context effect was strongest in stimuli with ambiguous phonetic prominence cues, the intonational effect, which is related to meaning, occurred independent of local phonetic prominence cues. There are at least two explanations for this different robustness. First, *meaning-based* effects as the one of the present study and the intonational one of [21] could affect directly the language processing (like the lexical-stress position) and make the listener more or less deaf for the corresponding phonetic cues. By contrast, other *pattern-based* effects like the rhythmic one of [21] could affect the interpretation of the local phonetic cues, which then provide the basis for language processing. If the context effect takes this way, clear phonetic cues leave less room for interpretation. Alternatively, the robustness of semantic-context effects could be matter of a more general cognitive strategy. As shown by many psycholinguistic studies (cf. [16]), speech perception is determined by statistical probabilities. This includes that more reliable cues outrank less reliable ones. Since the speech code serves to convey meanings and meaning relationships, it is logical that meaningful (e.g., semantic) units are more reliable predictors of other meaningful units than phonetic cues.

Follow-up studies should deal with these explanations in perception studies that address meaning- and pattern-based context effects beyond lexical stress and prominence. The perception studies should further be complemented by production studies, taking into account that the speaker is simultaneously a listener. From this perspective, if the speaker anticipates the signal-external cues that are available to the listener, the common presupposition that phonetic signal must always be self-contained may be questioned.

5. References

[1] Heid, S. & S. Hawkins (2000). An acoustical study of long-domain /r/ and /l/ coarticulation. *Proceedings of the 5th ISSP conference, Kloster Seeon, Germany*, 77-80.

[2] Wesener, T. (2001). Some non-sequential phenomena in German function words. *JIPA 31*, 17-27.

[3] Kohler, K.J. (1998). The disappearance of words in connected speech. *ZAS Working Papers in Linguistics 11, 21-34.*

[4] Niebuhr, O. (2009). The role of the vowel context in the differentiation of French /sʃ/ and /ʃs/ sequences. *Paper presented at the PaPI09 conference, Las Palmas, Spain.*

[5] Niebuhr, O. (2008). Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German. *JASA 142*, 1251-1263.

[6] West, P. (2000). Perception of distributed coarticulatory properties of English /l/ and /ɹ/. *J. of Phon. 27*, 405-426.

[7] Kuzla, C., H. Mitterer, M. Ernestus, & A. Cutler (2006). Perceptual compensation for voice assimilation in German fricatives. *Proc. 11th Australasian Int. Conf. Speech Science and Technology, Auckland, Australia*, 394-399.

[8] Smith, R. (2005). *The role of fine phonetic detail in word segmentation*. PhD Diss., University of Cambridge, UK.

[9] Ladefoged, P. & D.E. Broadbent (1957). Information conveyed by vowels. *JASA 29*, 98-104.

[10] Niebuhr, O. (2006). 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.

[11] Lindblom, B. (1990). Explaning phonetic variation. A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer.

[12] Niebuhr, O. (2008). Interpretation of pitch patterns and its effects on accentual prominence in German. *Paper presented at the 3rd TIE conference, Lisbon, Portugal.*

[13] Warren, R.M. & G.L. Sherman (1976). Phonemic restorations based on subsequent context. *Perception Psychophysics 16*, 150–156.

[14] Kouider, S. & E. Dupoux (2001). A functional disconnection between spoken and visual word recognition: evidence from unconscious priming. *Cognition 82*, 35-49.

[15] McGurk, H., & J. MacDonald (1976). Hearing Lips and seeing voices. *Nature 246*, 746-748.

[16] Clayards, M.A. (2008). *The ideal listener*. PhD thesis, University of Rochester, USA.

[17] Ladd, D.R. (1996). *Intonational phonology*. Cambridge: CUP.

[18] Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 126-152.

[19] Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech 21*, 347-353.

[20] Dilley, L. & J.D. McAuley (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language 59*, 291-311.

[21] Niebuhr, O. (2009). F0-based rhythm effects on the perception of syllable prominence. *Phonetica 66*, 95-112.

[22] Jensen, Ch. (2006). Are verbs less prominent? *Lund University, Centre for Language & Literature, Working Papers 52*, 73-76.

[23] Gussenhoven, C. & A.C.M. Rietveld (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics 16*, 355-369.

[24] Rosenvold, E. (1981). The role of intrinsic F0 and duration in the perception of stress. *ARIPUC 15*, 147-166.

[25] Kohler, K.J. (2005). The perception of prominence patterns. *Phonetica 65*, 257-269.

[26] Schneider, W., A. Eschman & Zuccolotto, A. (2001). E-Prime Reference Guide. Pittsburgh: Psychology Software Tools, Inc.