

Perception of Phonetic Detail in the Identification of Highly Reduced Words

Authors

Oliver Niebuhr^a

Klaus J. Kohler^b

Affiliation

^aDepartment of General and Comparative Linguistics, Christian-Albrechts-University, Kiel
Germany

^bInstitute of Phonetics and Digital Speech Processing, Christian-Albrechts-University, Kiel
Germany

Postal address of first author

Freyastrasse 10

24939 Flensburg

Germany

Phone : 0049 461 940 3993

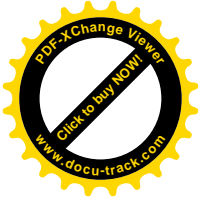
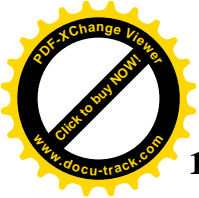
Email : niebuhr@linguistik.uni-kiel.de

Abstract

There is great phonetic variation of words in context, conditioned by phonetic environment, word type, and speaking style in different communicative situations. Function words and modal particles are particularly susceptible to having their phonetic weight and complexity reduced, especially in casual spontaneous speech. But even if whole strings of segments are no longer delimitable in reduced forms compared with fuller pronunciations of the same lexical items, there will still be *articulatory prosodies*, superimposed upon the remaining sound material, which retain essential components of the fuller forms, the *phonetic essence* that characterizes the whole form class of a word. The extreme reduction [aĩ] of the German modal particle *eigentlich* 'actually' [ai(g)ŋ.(t)(l)i(ç)] is a case in point. The length, palatality and nasality of its gliding movement reflect the polysyllabicity, the central nasal consonant and the final palatal syllable of the fuller forms. It is assumed that this phonetic essence triggers lexical identification in the listener. Therefore two perceptual identification experiments were carried out. They showed the crucial role of the duration of a palatal gliding section in the diphthong [aĩ] to distinguish between *eine* 'one' and *eigentlich* 'ne' 'actually a'. A third test showed further that listeners reacted differently to the palatal glide duration in different reduction environments, which may be related to different functional assessment of reduced forms in situational contexts.

Key words:

Speech reduction; Word identification; Speech perception, Perceptual restoration; Deletion; Phonetic detail; Articulatory prosody; Phonetic essence; German



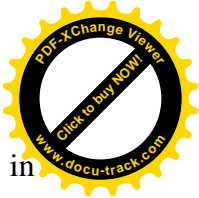
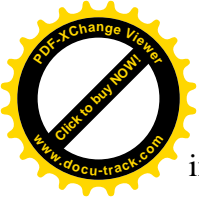
1. Introduction

1.1 The analysis of reduced speech

1
2 Words take on various phonetic manifestations in connected speech, depending on word type,
3 phonetic context, and speaking style, especially function words and modal particles. This is more
4 extreme in spontaneous dialogue. The traditional way of dealing with this phonetic variation is
5 to group the variants around canonical forms, which represent the most elaborate citation form
6 pronunciations, often guided by phonological transformation from orthographic form. This
7 canonical representation is generally segmental phonemic, and the variants are derived from it by
8 deletion, addition or modification of phonemic segments. If the changes, especially the
9 modifications, cannot be mapped one-to-one onto phonemic representations, allophonic
10 segmental statements are made at the phonetic level, e.g. when in *handbag* the realization of
11 /ndb/ is neither [ndb] nor [mb] but [n̠mb] with coronal-labial double articulation, or when the
12 assimilated fricative in *this* [ʃʃ̚] *shop* is different from the geminate fricative in *fish* [ʃʃ̚] *shop*.
13 This segmental representation of words in relation to phonemic canonical forms is a useful
14 sorting principle for pronunciation dictionaries, such as the reference works by Jones/Roach
15 (1997) and Wells (1990) for English, and WDA – Wörterbuch der deutschen Aussprache (1969)
16 for German.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 The canonical-phonemic reference approach to phonetic variation of lexical entries is also a
33 useful heuristic device for systematic descriptions of the pronunciation of a language, such as
34 Gimson/Cruttenden (1962, 2008) for English, or Kohler (1995) for German, as well as for
35 descriptive models of speech reduction, e.g. Kohler (1990, 1998, 2001) for German. Finally, the
36 segmentation and labelling of acoustic speech corpora has greatly profited from this framework.
37 However, it runs into conceptual problems when the distinctive features of vowels and
38 consonants as well as their assimilation or elision are no longer linearly segmentable (cf. Nolan
39 1992; Gow 2002; Local 2003; Heid and Hawkins 2000), and when phoneme strings, which may
40 extend beyond syllables to whole words, need to be marked as deleted qua segmental units
41 although the signal portion is still recognized as containing the full lexical information in the
42 utterance context.

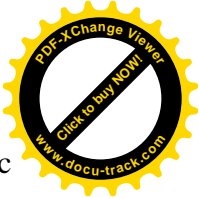
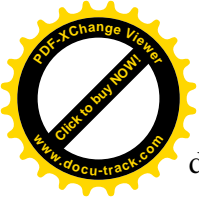
43
44
45
46
47
48
49
50
51
52
53
54 As regards the latter case, the following example of highly reduced speech was found in the Kiel
55 Corpus of Spontaneous Speech (IPDS 1995, 1996): *ich kann Ihnen das ja mal sagen* ‘I can
56 mention this to you’ (g072a015). Figure 1 provides the speech wave, the spectrogram, as well as
57 a linear segmentation and SAMPA labelling. There is a long stretch of an alveolar nasal (180 ms)
58
59
60
61
62
63
64
65



in [ka__as] (cf. dotted box in Fig.1), i.e. from the end of the word *kann* into the word *das*, in which there is initial nasalization instead of a lenis plosive. It is not possible to delimit the word *Ihnen*, whose citation form phonemic transcription is /i:nən/. Its usual pronunciation is without [ə] in the second syllable, but as is obvious from Figure 1, the utterance shows no delimitable signal portion for a vowel [i:] either. However, the person who did the orthographic transliteration, and subsequently the labeller as well as the phoneticians who processed the data, had absolutely no doubt that the utterance contained *Ihnen*, and was not *ich kann das ja mal sagen*, without *Ihnen* but also with a long [n] in [ka__as], an equally possible utterance in the same semantic context of an appointment-making scenario. Looking more closely at the phonetic manifestation of the nasal stretch reveals that it has palatalization throughout, increasing towards the centre. Kohler (1999) referred to such articulatory residues in the reduction of function words as *articulatory prosodies*, “which persist as non-linear, suprasegmental features of syllables, reflecting, e.g., nasality or labiality that is no longer tied to specific segmental units” (p. 89). Thus, articulatory prosodies are distinctive suprasegmental vocal-tract and phonation features that identify words in context in spite of segmental reduction. In the above example, the articulatory prosody of palatalization, in addition to the long duration of the signal portion of the alveolar nasal, which it overlays, references *Ihnen*.

//////////INSERT FIGURE 1 ABOUT HERE//////////

Of course, what happens here from an articulatory point of view is that the tongue body articulation for the high front vowel [i:] in between the open central vowels of [ka__as] is not elided but is carried out while the tongue tip/blade forms contact with the alveolar ridge, resulting in palatalization rather than in an acoustic [i:] segment as defined by phoneticians. From this perspective, the weakly reduced form [i:n̠] and the more strongly reduced form [n̠n̠] can be related to the same class (i.e. *Ihnen*) without an elaborate derivation from one canonical representation, because they both contain palatality and long alveolar nasality, as do other intermediate degrees of reduction. This means that all phonetic forms of this word must contain these features; they constitute the *phonetic essence* of *Ihnen*. This concept of phonetic essence may be assumed to apply to function words generally and possibly even to all lexical items. The phonetic essence of a lexical item manifests itself either in segmental units in the less reduced forms or as articulatory prosodies in more extreme reduction, where it appears to be sufficient for the listener to identify the word. Thus, [kan̠n̠as], as against [kannas] *kann das*, can be

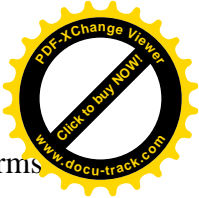
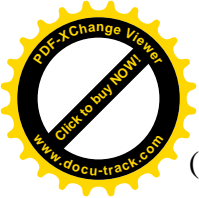


1 decoded as containing *Ihnen* although there is no syntactic/semantic bias in the same linguistic
2 and situational context.

3
4 Contrary to the conventions guiding vowel and consonant segmentation, such articulatory
5 prosodies are not temporally delimited; all that matters is that they manifest themselves within a
6 certain environment, where exactly can vary greatly. The extension may actually be quite large,
7 as in the present case where palatality reaches beyond the nasal consonant in both directions: the
8 aspirated plosive [k^h] of the preceding word *kann* is fronted, and its vowel as well as the one of
9 the adjacent word *das* are raised and centralized. In *kann leider* of another utterance by the same
10 speaker, where *kann* is also followed by an alveolar consonant and in turn by an open vowel
11 element, this is not the case.
12
13
14
15
16
17
18
19

20
21 Glottalization is another articulatory prosody in German, used to signal interruption of modal
22 voice in nasal context in place of a stop consonant. Thus, *könnten* ‘could’ may be differentiated
23 from *können* ‘can’ [kœnn], either as [kœntn] or [kœnn̥], both creating the same kind of “break”
24 in the acoustic signal for the listener. In the case of [t], the break is referable to a segmentally
25 delimited stop occlusion inside a stretch of alveolar nasality; in the case of [ŋ], it is superimposed
26 on the nasal stretch and variable as to its extension and position. As long as a few irregular glottal
27 pulses occur somewhere in the signal portion of [œnn̥] the listener decodes it as the lexical unit
28 *könnten*, cf. Kohler (1999). The same articulatory prosody of glottalization can differentiate
29 German *campen* ‘to camp’ [k^hɛmm̥] from *kämmen* ‘to comb’ [k^hɛmm] and ‘*sollten* ‘should’
30 [zɔl̥n] from *sollen* ‘are to’ [zɔln], for example, in the contexts *Zum campen/kämmen ist es noch*
31 *zu früh* ‘It is still too early to camp/comb’ and *Sie sollten/sollen das machen* ‘They should/are to
32 do this’. Again, glottalization produces a signal break, as does the stop articulation of [t] and [p].
33 This break is part of the phonetic essence of such words as *könnten*, *campen* and *sollten*, in
34 distinctive contrast to *können*, *kämmen* and *sollen*.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50
51 A third type of articulatory prosody in German is nasalization. Strongly reduced forms of *soll er*
52 and *sollen wir* in the frame ___*das machen?* ‘is he / are we to do this?’ may be contrasted as
53 [zɔə] versus [zɔmə]. In *sollen wir*, the bilabial occlusion may not be reached but lowering of the
54 velum still occurs, thus nasalization is superimposed on the vocalic resonances – [zɔ̃ə], replacing
55 a segmentally delimitable nasal consonant by an articulatory prosody with the same function of
56 signalling the phonetic essence of nasality to a listener, which is absent from *soll er*, cf. Kohler
57
58
59
60
61
62
63
64
65



(1999). In other languages like English and French, observations which are interpretable in terms of articulatory prosodies concern, for example, long-term resonances of light and dark liquids (cf. Heid and Hawkins 2000) or quality, f₀ and duration changes of consonants in the context of schwa elision (cf. Gadet 1992).

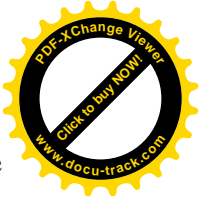
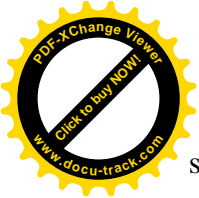
The segmentation and labelling conventions that were developed for the phonetic annotation of the Kiel Corpus took articulatory prosodies for lexical identification into account. If phonemic segments of a canonical base form of a lexical item cannot be mapped onto a signal portion corresponding to it they have to be marked as deleted, but if at the same time there is still an articulatory prosody that carries lexical identification it needs to be marked as a non-segmental insertion <-MA> (= prosodic marker) put before the segment string marked as deleted. So, *Ihnen das* in Figure 1 is transcribed as

<##-MA \$Q- \$i:- \$n \$@- n-+ ##%d-n \$a \$s+>

(for further details of the transcription system and conventions cf. Kohler et al. 1995). Search operations on corpora annotated in this way can access the incidence of articulatory prosodies, and the general prosodic marker <-MA>, which only references their occurrence, without specifying their phonetic manifestations, can then be given a detailed phonetic analysis. Thus the *Kiel Corpus* annotation continues to use a canonical-phonemic framework as a heuristic device and enriches it by articulatory prosodies to deal more satisfactorily with the systematic distinctive phonetic representation of lexical items across their contextual and situational variability. But the addition of this symbolization of non-segmental articulatory attributes is still part of a wider heuristics. It does not provide a language model, which can only follow from the subsequent analysis of data preprocessed, and thus made accessible, in this way. The imperfections of any manual phonetic labelling system of acoustic data, with regard to reliability, remain. This approach incorporates ideas of Firthian prosodic analysis (Firth 1948) in the phonetic description of a language and proposes the concept of complementary phonology (Kohler 1994).

1.2 Lexical access and phonemic restoration

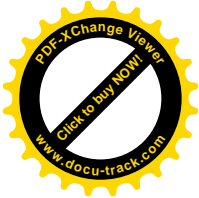
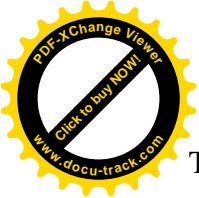
A number of studies on the recognition of reduced word forms have assumed that the mental lexicon contains representations only for the canonical forms of lexical items and that lexical access of reduced forms involves a process of reconstructing the canonical form (cf. Gaskell and Marslen-Wilson 1996, 2001; Gow 2002; Mitterer and Blomert 2003 for discussions of this issue). Initially such studies focussed on the lower end of speech reduction. For example, a word-final /n/ that is assimilated towards [m] in utterances like *gun production* can differ in its spectral details from an actual /m/ as in *gum production* (Gow 2002; cf. also Holst and Nolan 1996 for



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

silabiant assimilation). Even if the assimilation of the sound itself is complete and the /n/ in the above example becomes indistinguishable from /m/, the assimilated sound can still leave traces in the preceding vowel. Such subtle phonetic detail allows listeners to identify the assimilated or deleted sound and hence to differentiate between utterance pairs like *gun production* with /n/ → [m] and *gum production* (Gow 2002; cf. also Nolan 1992; Manuel 1992; Niebuhr et al. 2009 for similar observations in other assimilatory processes). This ability or readiness of listeners is conditioned by the context of the surrounding segments, which trigger the processes of assimilations or deletions, and the detailed phonetic exponents are then treated as mediating the reconstruction process (cf. Gaskell and Marslen-Wilson 1996, 2001; Mitterer and Blomert 2003; Snoeren et al 2006; Mitterer and Ernestus 2006; Mitterer et al. 2008).

At higher levels of reduction, particularly in everyday spontaneous communication, as illustrated for German in 1.1, entire strings of segments may be “missing”. Another well investigated example is the Dutch word *eigenlijk* ‘actually’ [ɛiχəɫək^h], which can be realized as [ɛɛg^ɪ]. The perception experiments of Ernestus et al. (2002) and Kemps et al. (2004) showed that at this higher level of reduction signal-external factors were more relevant for word identification than the directly surrounding phonetic context (cf. also Gaskell and Marslen-Wilson 2001). Kemps et al. (2004) carried out experiments on Dutch words ending in the suffix *-lijk*. A non-reduced realisation with [l], and a reduced one without, of each selected word were excerpted from a corpus of spontaneous speech and presented either with a full context or minimally as the suffix, by itself or also including preceding sound portions. Participants had to perform a phoneme-monitoring task on [l]. In a significant number of instances, listeners reported the presence of [l] in the full context, although it was not in the signal they heard. This rarely happened in the minimal context. The authors interpret this finding as showing that “Full context [...] provides listeners not only with all phonetic cues but also with syntactic and semantic information.” (Ernestus et al. 2002:169), and that “listeners restore phonemes that are missing in reduced word forms” (Kemps et al. 2004:120). The segmental and semantic prerequisites of this approach argue against the possibility that an articulatory prosody representing a phonetic essence can play a crucial role in word identification. Rather, they suggest a perceptual process in which the phonetic input is traced back to the richer and more abstract representation of the canonical form, and which thus tries to recover a (highly) reduced word on the basis of its remaining segments with support from the syntactic and semantic context.



Therefore, the goal of our study is to provide an answer as to whether subjects can perceive highly reduced word forms for lexical identification, in the same syntactic and semantic frame, when these forms contrast in the presence or absence of articulatory prosodies, as defined in 1.1.

1.3 The phonetics of German *eigentlich*

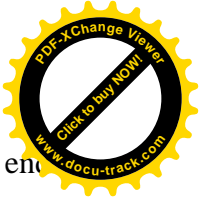
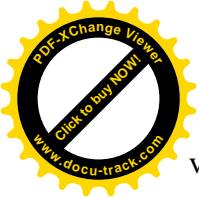
The perception experiments to be reported in this paper take their point of departure from the German modal particle *eigentlich*, which in its broad array and extreme degree of reduction is parallel to the Dutch particle *eigenlijk*, cf. Ernestus (2000). The stimuli of the perception experiments build on production data that came from a database search of the Kiel Corpus of Spontaneous Speech and were reported in Kohler (2001). The phonetic variation of *eigentlich* may be summarized under three headings:

- the first 2 syllables are commonly [aɪŋŋ]
- the last syllable, [liç], has a closer vowel than, e.g., the female suffix *-in*, and is characterized by an articulatory prosody of palatality throughout
- the juncture between these word portions may be [t] or assimilated [k], or it may be absent.

In these sections of *eigentlich* further reductions occur. The form [aɪŋŋ] may be reduced to [aŋ], a case of nasalization of lenis plosives before nasals, found in all German words ending in unstressed *-gen*. In [liç], the transition from an alveolar lateral to a high front vowel involves a complete reversal from front contact and side back opening to front opening and side back contact, which is frequently resolved by giving up the lateral articulation. Furthermore, in the unstressed position airflow may be reduced to such an extent that the fricative noise no longer surfaces. These articulatory conditions result in the most reduced form [aɪŋi] found in corpus.

As explained for [zɔmɐ] > [zõẽ] in 1.1, [aɪŋi] may be further reduced to [aĩ]. It is bisyllabic or contains a long gliding to a high front vowel position. Additionally, it is characterized by nasalization across this gliding portion. This extremely reduced form thus retains the palatality and the nasality as well as a duration feature of the fuller forms. The three components palatality, nasality, and duration constitute the phonetic essence of the class of reduced forms of German *eigentlich*, and they are still present as articulatory prosodies in the most reduced one.

These phonetic descriptions have been complemented by a more detailed acoustic analysis of the 56 *eigentlich* tokens that were uttered by 10 female and 9 male speakers in the Kiel Corpus. The



vocalic portion, inside the initial diphthong, from the onset of the upward F2 movement to its end was set in relation to the total word duration. This analysis showed that the more the word is reduced the more of the palatality that characterizes the final syllable is transferred into the gliding section in the initial diphthong, which also represents a palatal gesture. First, there is a trend (Pearson, $r=-0.29$, $n=54$, $p<0.1$) for shorter *eigentlich* productions to have higher second-formant frequencies at the end of the glide. A higher second-formant frequency indicates a closer approximation of the palatal region by the tongue body, i.e. the end of the glide becomes more [i]-like. More importantly, however, the enhancement of palatality for greater degrees of reduction showed up in a temporal reorganization of the diphthong in favour of a lengthening of the gliding section. As is illustrated in Figure 2, there was a highly significant negative correlation (Pearson, $r=-0.48$, $n=54$, $p<0.001$) between the overall word duration and the proportion of the gliding section in the diphthong.

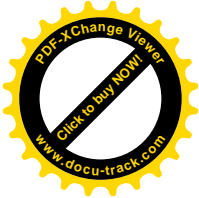
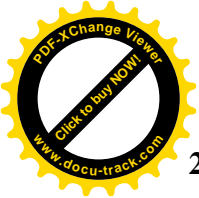
//////////INSERT FIGURE 2 ABOUT HERE//////////

1.4 Devising a perception experiment

With reference to the possible reduction [ãĩ], two utterances may be constructed:

- *eine rote* ‘a single red one’, *eine* realized as unreduced [ainə], with contextual nasalization in the diphthong, lack of reduction signals accentuation;
- *eigentlich* ‘*ne rote* ‘a red one, really’, ‘*ne*’ [nə] being a reduced form of *eine* and *eigentlich* realized as [ãĩ].

According to 1.3 the two utterances differ in the duration of the palatal gliding section, and the height of its endpoint, which, given sufficient duration, would be less crucial. They correspond in the nasalization of the gliding section, which is contextual in *eine* but an essential articulatory residue in reduced *eigentlich*. So, a perception experiment can now set out to investigate the influence on word identification of degrees of palatality in gliding movements from an open vowel. The rationale for such an experiment is to start from a natural production of *eine rote*, as defined above, and to modify the durations of the open vowel portion and the palatal gliding section independently. The hypothesis is that the presence of long palatality will trigger *eigentlich* ‘*ne*’ judgements, with a critical duration value for the change-over from *eine*; lengthening the open vowel, on the other hand, will always result in *eine*, with varying degrees of accentual prominence.



2. Perception Experiments

2.1 General

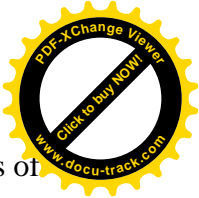
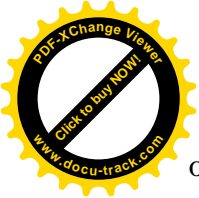
1
2 The hypothesis underlying the perception experiments is that perception reflects the production
3 of the extreme reduction form [aĩ] of the lexical item *eigentlich*, i.e. a long duration of a palatal
4 gliding section, on which nasalization is superimposed. The acoustic analysis of the corpus data
5 has shown that the duration of this portion increases by at least 60 ms from fuller forms to
6 strongly reduced forms of *eigentlich* (cf. 1.3). This value is therefore considered perceptually
7 critical, and the change-over in perception from *eine* to *eigentlich* 'ne is expected to occur at this
8 order of magnitude in the palatal gliding section. This expectation is further supported by the fact
9 that 60ms should be well above the just noticeable difference for duration changes in speech
10 stimuli, cf. Lehiste (1970), Klatt and Cooper (1975).

11
12
13
14
15
16
17
18
19
20
21 Two perception experiments were run aiming at direct and indirect identification, respectively, of
22 *eine rote* or *eigentlich* 'ne rote. Their purpose was to determine whether and to what extent an
23 articulatory prosody of long palatality, as a component of the phonetic essence of reduced word
24 forms of *eigentlich*, affects perception and lexical identification. In an addition to the indirect
25 identification experiment, the same test stimuli were presented in the same verbal context, which
26 was, however, pronounced with a different degree of phonetic reduction. This design is to test
27 whether listeners become aware of degrees of reduction and then react differently to test stimuli
28 because of a different functional assessment in different reduction environments. The test stimuli
29 for all the experiments were taken from the same generated series.

30
31
32
33
34
35
36
37
38
39 The experiments exclude semantics as a separate factor. Of course, lexical decoding is embedded
40 in a general semantic frame, but it does not favour one or the other interpretation of the target
41 stimuli.

2.2 Test stimulus generation

42
43
44
45
46
47 The stimulus generation was done in three steps. First, the short utterance *eine rote* was produced
48 naturally by a trained phonetician, the first author (ON), in a moderately reduced fashion. The
49 two disyllabic words were realized as comparably salient with an overall roughly flat f0 course.
50 That is, the two words were stressed, but not accented on the initial syllables (cf. Ladd 1996).
51 The natural production of *eine rote* was the starting point for two stimulus series that were
52 created in the second step using the PSOLA resynthesis of Praat, cf. Boersma (2001). As shown
53 in Figure 3, the stimuli of the two series resulted from duration manipulations in 100ms sections
54 that covered either [ɪn^j] or [ə] of *eine*. In one series, the duration of the [ə] section retained its



original 100ms duration, while the [ɪn^j] section was lengthened linearly in 6 equal-sized steps of 20ms (i.e. 120ms, 140ms, 160ms, etc.). A stimulus was resynthesized for each step. In the other series, a parallel lengthening was done for the [ə] section, while the [ɪn^j] section was kept constant. Again, a stimulus was resynthesized after each lengthening. Hence, the two stimulus series consist of 7 stimuli each. In both series stimulus 1 is the naturally produced *eine rote* with the original durations of the [ɪn^j] and [ə] sections, whereas in stimulus 7 the respective section is 120ms (or 120%) longer than in the original production (i.e. 220ms). Indicating the lengthened section, the two stimulus series will be referred to as the EINE_{in+} and EINE_{a+} series.

//////////INSERT FIGURE 3 ABOUT HERE//////////

In the third step, a further round of PSOLA resyntheses replaced, in all 2x7=14 stimuli, the flat f0 course with a phonologically constant intonation pattern that consisted of two rising-falling pitch-accent peaks on the syllables *ei-* and *ro-*, the latter ending in a terminal, low f0 level at the stimulus offset. The first pitch peak on *ei-* had a range of 6 semitones. The second pitch peak on *ro-* was downstepped and hence 1 semitone lower than the first peak. The f0 rises and falls in between the onsets, maxima, and offsets of the peaks were linearly interpolated. The frequency values of the onsets, maxima, and offsets remained constant across all stimuli. However the temporal positions of the peak maxima were adjusted to the duration manipulations, i.e. they were placed at a fixed distance of 60ms after the vowel onset of the corresponding syllable. The pitch accent on *eine* allows its interpretation as a numeral ‘a single’, instead of as an indefinite article.

2.3 Experiment 1: direct identification test

2.3.1 Method

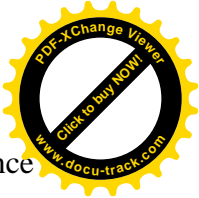
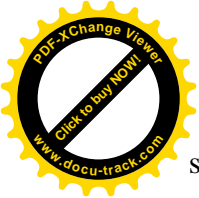
Stimuli

Experiment 1 is based on a subset of the generated stimuli, including stimulus 1, i.e. the naturally produced *eine rote* with the original duration structure, as well as the two stimuli 7 of the EINE_{in+} and EINE_{a+} series, which show the greatest lengthening of the [ɪn^j] and [ə] sections.

Procedure and participants

The three stimuli were integrated into a larger dictation task, in order to make it impossible for the listeners to uncover the aim and the target wordings of the listening test. For this purpose, 10 short conversational texts were created. Three of them provided the framework for the stimulus

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



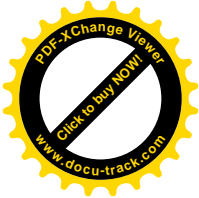
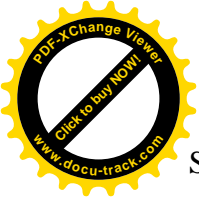
subset. That is, each text contained one of the stimuli as a separate syntactically elliptic sentence in the second turn of the conversation. The semantic content of each of the three texts was basically compatible with the interpretation of the stimuli both as *eine rote* and as *eigentlich 'ne rote*. Text (A) deals with a conversation during a poker game. In texts (B) and (C) the stimuli occurred in the contexts of a political election or a wine tasting. The three experimental texts are given in the Appendix.

In addition to these three experimental texts, 7 distraction texts were included in the dictation task. They had a similar make-up as the three experimental texts with regard to the overall length and the internal structure, i.e. an introductory sentence was followed by a single two-turn dialogue. All 10 conversational texts were read in an equally casual fashion by the first author (ON) and recorded digitally in a sound-treated booth at the Institute of Phonetics and Digital Speech Processing (IPDS) of the University of Kiel. In reading the three experimental texts, the stimulus slot was filled with a reduced variant of *eigentlich 'ne rote*. In the subsequent post-processing, these filler variants were replaced by each of the three stimuli. This complete cross-combination yielded $3 \times 3 = 9$ versions of the experimental texts.

The experiment was done with a group of 45 undergraduates as part of a course on Spontaneous Speech at the Department of General and Comparative Linguistics of the University of Kiel. The 45 students were all native speakers of German with no known hearing disorders. They were divided into three subgroups of 15 subjects, with gender and age approximately balanced across the subgroups. Two of the groups consisted of 10 female and 5 male subjects with average ages of 23.4 or 24.1 years. The remaining group consisted of 9 female and 6 male subjects. They were on average 23.7 years old.

Each of the three subgroups did the experiment in a separate session. Each session contained the same 7 distraction texts and different versions of the three experimental texts. That is, in the three sessions the three stimuli were framed by different texts. In this way, 15 responses were collected across the 9 combinations of experimental texts and stimuli, although each subject heard each experimental text and each stimulus just once. Moreover, the 7 distraction texts and the three experimental texts were arranged in an overall differently randomized order for each session.

At the beginning of each session, the subjects were informed orally that they were to do 10 short dictations of casually read conversational texts as part of a class dealing with the relationship between phrasing, turn yielding, and punctuation. This pretext was plausible in a course on



Spontaneous Speech, and the reference to punctuation further distracted the subjects from the actual aim of the experiment, i.e. the wording of the stimuli. The subjects were then asked to write down what they heard, using appropriate punctuation. Each of the 10 short dictations in a session was done in the same way. First, the conversational text was played as a whole. Then, each of the three elements of the text, i.e. the introductory sentence, the first turn, and the second turn, were played separately and repeated several (mostly three) times, until all subjects finished writing down what they heard. Then, the next dictation started. At the end of the session, the written texts were collected for subsequent analysis with regard to the wordings of the stimuli.

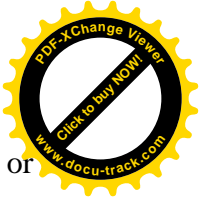
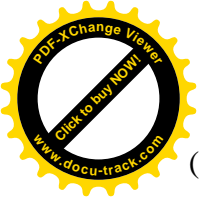
The conversational texts or text elements were played by means of Adobe Audition (<http://www.adobe.com/products/audition/>) with a constant loudness, and presented to the listeners over loudspeaker in a quiet lecture room at the University of Kiel. The separate dictation sessions with the three subgroups took about 45-60 minutes each.

2.3.2 Results and discussion

The analysis of the written dictations showed that all 45 listeners of the three subgroups transliterated the test stimuli in the written texts as either *eine rote* or *eigentlich 'ne rote*. The specific absolute and relative frequencies of *eigentlich* occurrences within and across the subgroups (white columns vs. grey column) are summarized in Table 1. It is immediately obvious that stimulus 7 of the EINE_{in+} series was almost exclusively transliterated as *eigentlich 'ne rote*. Just one of the 3x15 listeners transliterated the stimulus as *eine rote*. Hence, the overall *eigentlich* identification triggered by stimulus 7 of the EINE_{in+} series amounts to 97.8%. By contrast, less than 25% of the 3x15 listeners identified and transliterated stimulus 7 of the EINE_{a+} series as *eigentlich 'ne rote*. The common starting point of the EINE_{in+} and EINE_{a+} series, i.e. stimulus 1 with the original duration structure, evoked the word *eigentlich* in only 5 of the 3x15 transliterations or in 11.1% of 45 cases.

//////////INSERT TABLE 1 ABOUT HERE//////////

The results were analyzed, parallel to the structure of Table 1, in a 3x3 contingency table by two three-level factors, viz. stimulus (cf. rows in Table 1) and experimental text (cf. white columns in Table 1). Since the listeners of Experiment 1 produced single transliterations (i.e. categorical, nominal decisions) for each stimulus, non-parametrical χ^2 tests were done in order to test whether the *eigentlich* frequencies were significantly differently distributed across the levels of the two factors. We did not calculate individual χ^2 tests for the three text conditions with transliteration



(*eigentlich* vs. *eine*) as a separate two-level factor. Since every listener wrote either *eigentlich* or *eine* for each stimulus, the combined *eigentlich* and *eine* frequencies of each cell in Table 1 sum up to 15. The *eigentlich* and *eine* frequencies per cell can be transformed into each other, so it is sufficient to base the χ^2 tests just on the *eigentlich* frequencies.

As was expected from the descriptive statistics, the variable stimulus had a highly significant effect on the cell frequencies ($\chi^2=15.05$; $df=4$; $p<0.01$). However, even though there are indications in Table 1 that *eigentlich* was identified and transliterated less frequently in the poker text than in the wine text, the variable ‘experimental text’ had no separate significant effect ($\chi^2=1.08$; $df=4$; $p>0.05$). This outcome accords with experimental texts that are semantically compatible with both the *eine rote* and the *eigentlich ‘ne rote* readings of the stimuli.

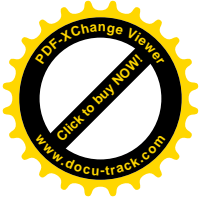
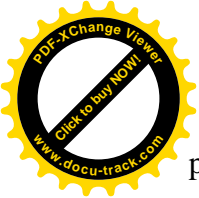
2.4 Experiment 2: indirect identification test

2.4.1 Preliminaries

The results of Experiment 1 show that listeners can confidently identify test stimuli as either *eine rote* or *eigentlich ‘ne rote* on the basis of the make-up of an [a]-to-[ɪ] movement, in situational semantic contexts that allow either interpretation. This shows that listeners can detect articulatory prosodies in highly reduced speech and relate them to the phonetic essence of a word. It needs to be determined now what the critical duration values of the [a]-to-[ɪ] movement are for a change-over from *eine rote* to *eigentlich ‘ne rote* judgements to occur. For this purpose the design of an indirect identification test is used, which puts the test stimuli in a constant context, and subjects are asked whether context and test stimulus match semantically.

So far, indirect identification tests have primarily been used to determine intonational categories and their abstract meanings (cf. Nash and Mulac 1980; Kohler 1987; Kleber 2006; Niebuhr 2007). An advantage of this method is that it does not require judging stimuli explicitly with regard to specific (meta)linguistic labels, such as the wording of the utterance or particular sounds or phonemes. Instead, the listeners make indirect statements about such labels by judging whether the stimuli do or do not match with a constant preceding context utterance. In this way, perceptual effects become observable that are difficult to ascertain by simple (meta)linguistic labels, and listeners are not alerted to these labels or to the aim of the experiment.

In the present case, the match is to be one of word semantics. The context utterance is the question *wieviele willst Du?* ‘how many do you want?’. It was realized by a female Standard German speaker in a casual way with an intonation contour consisting of prenuclear and nuclear

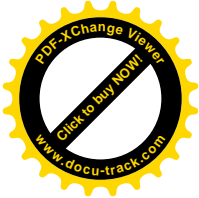
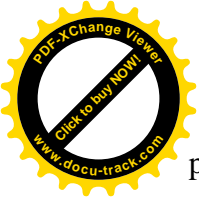


1 pitch accents that rise across the syllables *wie-* and *willst*. The prenuclear and nuclear pitch
2 accents are concatenated by a f0 fall, and the nuclear accent leads over to a high rise until the end
3 of the question utterance. It is obvious that a semantically sensible and hence matching answer to
4 the question *wieviele willst Du?* can only be the utterance *eine rote* ‘a single red one’, in which
5 the accented *eine* represents a numeral; *eigentlich* ‘*ne rote* ‘a red one, really’ does not convey
6 numerical information and therefore does not match the question precursor. Consequently, by
7 judging the question - stimulus pairs as matching or not matching the listeners provide indirect
8 information about the wording they identified in the stimuli. Admittedly, a non-matching
9 response does not *per se* imply that the listeners perceived *eigentlich* ‘*ne rote*. However, since the
10 EINE_{in+} stimulus of Experiment 1 was uniquely perceived as *eigentlich* ‘*ne rote* and is to occur
11 again as the endpoint of a 7-point duration scale for the palatal gliding section in Experiment 2,
12 there is strong reason to assume that non-matching judgements do not simply exclude *eine rote*
13 but positively identify the perception of *eigentlich* ‘*ne rote*. It is further predicted on the basis of
14 1.4 that transition from matching to not matching occurs for the EINE_{in+} series, but not for the
15 EINE_{a+} series.
16
17
18
19
20
21
22
23
24
25
26

27 2.4.2 Procedure and participants

28 Fourteen question - stimulus pairs were created by attaching the 2x7 stimuli of the EINE_{in+} and
29 EINE_{a+} series of 2.2 to the casual w-question *wieviele willst Du?*. Question and stimuli were
30 separated by 350ms, which is cross-linguistically a typical silent interval for a non-overlapping
31 turn change, cf. Weilhammer and Rabold (2003). The perception experiment contained 10 copies
32 of the 14 question - stimulus pairs. They were arranged in an overall randomized order and
33 divided into 14 blocks of 10 pairs, following the organization of the answer sheets. Blocks were
34 separated by double beeps, question - stimulus pairs by single beeps. Each question - stimulus
35 pair was followed by a pause of 4 seconds during which the listeners had to make their
36 judgements.
37
38
39
40
41
42
43
44
45
46

47 The listeners were 20 native speakers of German (12 females and 8 males, average age 23.8
48 years) with no known hearing disorders. They were all undergraduates at the Department of
49 General and Comparative Linguistics of the University of Kiel and naïve with regard to the
50 experimental background. At the beginning of the experiment, the listeners received written
51 instructions that were simultaneously played over loudspeakers and also included examples
52 illustrating the phenomenon of speech reduction. They were informed that they would hear
53 question - answer pairs, and that they were to judge whether the answer did or did not match the
54
55
56
57
58
59
60
61
62
63
64
65



preceding question. Decisions were to be made spontaneously by ticking boxes on prepared answer sheets.

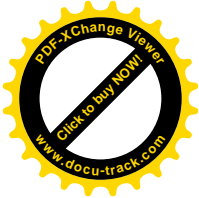
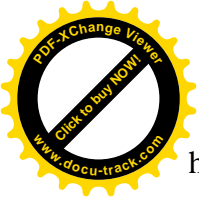
After the instructions, the listeners judged a block of 12 question - stimulus pairs, which familiarized them with the procedure and the nature of the stimuli. This practice block consisted of the pairs that contained the stimuli with the greatest lengthening (i.e. stimulus 7), with the middle lengthening (i.e. stimulus 4) and with the original duration (i.e. stimulus 1) from both the EINE_{in+} and the EINE_{a+} series. The 3x2=6 pairs were presented twice in an overall randomized order. The entire experiment took around 35 minutes and was done in a quiet and sound-treated room at the IPDS in Kiel. The question - stimulus pairs were presented via loudspeakers.

2.4.3 Results and discussion

A total of 200 judgements were obtained for each stimulus. Figure 4 summarizes the total sums across the stimulus series in terms of percentages of not-matching judgements. The results of Experiment 2 were also analyzed by means of a two-factor repeated measures ANOVA based on the independent within-subject factors type of lengthened section, i.e. [ɪn^j] vs. [ə], and degree of lengthening, which corresponded to the stimulus numbers 1-7 in the EINE_{in+} and EINE_{a+} series. Higher stimulus numbers represent greater lengthening. The dependent variable is the sums of not-matching judgements that were produced by the 20 individual listeners for each stimulus across all 10 repetitions. Thus, each of the 20 measurements in the conditions created by the two within-subject variables varies between 0-10. When a factor in the repeated-measures ANOVA violated the assumption of sphericity, as determined by Mauchly's test, the Greenhouse-Geisser correction was applied. In reporting the results of that factor, we provide the corrected p level, but the original, uncorrected *df*'s. All calculations were done in SPSS (Landau and Everitt 2004).

//////////INSERT FIGURE 4 ABOUT HERE//////////

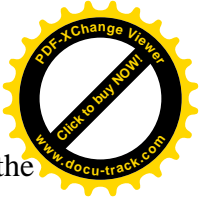
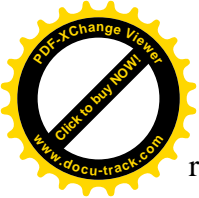
Figure 4 shows that the grey curve, which represents the EINE_{in+} series, is characterized by a rapid transition from matching to not matching across stimuli 1-7. While stimulus 1 evoked less than 10% not-matching judgements, stimuli 5-7 were judged almost 100% as not matching. A lengthening-related increase in not-matching judgements can also be observed for the EINE_{a+} series. However, the increase stagnates at around 30% for stimuli 5-7, which is just around the not-matching level already reached by stimulus 2 of the EINE_{in+} series. The fact that both the EINE_{in+} and the EINE_{a+} series yielded more not-matching judgements with higher stimulus numbers, but to a different extent, also manifested itself in the output of the ANOVA. It yielded



1 highly significant main effects of both within-subject factors (type of lengthened section:
2 $F_{(1,19)}=1,008.207$; $p<0.001$; $\eta_p^2=0.982$; degree of lengthening: $F_{(6,114)}=155.020$; $p<0.001$;
3 $\eta_p^2=0.891$) as well as a highly significant interaction between them ($F_{(6,114)}=71.903$; $p<0.001$;
4 $\eta_p^2=0.719$). All significances are linked to considerable effect sizes in terms of partial eta-
5 squared (η_p^2).
6
7
8
9

10 On this basis, we split the factor ‘type of lengthened section’ and ran two additional repeated-
11 measures ANOVAs that addressed each of the stimulus series separately. In the ANOVA for the
12 $EINE_{in+}$ series the degree of lengthening became highly significant ($F_{(6,114)}=331.907$; $p<0.001$;
13 $\eta_p^2=0.993$). The effect size of $\eta_p^2=0.993$ indicates that this factor was able to explain virtually all
14 variance in the not-matching judgements. The ANOVA of the $EINE_{a+}$ series also showed a
15 significant main effect of the degree of lengthening ($F_{(6,114)}=18.461$; $p<0.001$; $\eta_p^2=0.688$), but the
16 significance level (cf. F statistics) and the effect size were substantially lower than in the case of
17 the $EINE_{in+}$ series. Furthermore, the ANOVA for each stimulus series included all possible
18 pairwise post hoc comparisons between the 7 levels of the factor ‘degree of lengthening’
19 ($7 \times 6 = 42$ post hoc tests with Bonferroni corrections integrated into p levels; thus the threshold for
20 the corrected significances remained at $p<0.05$). We restrict the report of these post hoc results to
21 the pairs of adjacent stimuli that represent the progression of lengthening across the $EINE_{in+}$ and
22 $EINE_{a+}$ series. As the judgement curves in Figure 4 either rise or run roughly flat, all other
23 significances can be inferred from the reported ones. As regards the $EINE_{in+}$ series, the not-
24 matching judgements increased significantly in stimuli 1-2 ($p=0.009$), 2-3 ($p<0.001$), 3-4
25 ($p=0.033$), and 4-5 ($p<0.001$). The further lengthening of the [ɪnʲ] section across stimuli 5, 6, and
26 7 did not cause further significant increases in not-matching judgements. The successive [ə]
27 lengthening across the $EINE_{a+}$ series resulted in only one significant increase in not-matching
28 judgements that occurred in stimuli 4-5 ($p<0.001$).
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 Overall, the results show that the stimuli produced systematic response patterns. The listeners
48 were not guessing and made their matching – not-matching decisions with regard to a consistent
49 perceptual measure. Seen in relation to the transliterations of the stimuli in Experiment 1, it is
50 reasonable to assume that this consistent perceptual measure was the *eigentlich* identification in
51 the stimuli. On this basis, Experiment 2 provides further evidence for the conclusions drawn
52 from Experiment 1. Adding features of the phonetic essence of *eigentlich*, duration and
53 palatality, to the stimuli by lengthening the [ɪnʲ] section across the $EINE_{in+}$ series induced a clear
54 reinterpretation of the stimulus wording from *eine rote* to *eigentlich 'ne rote*. Such a
55
56
57
58
59
60
61
62
63
64
65



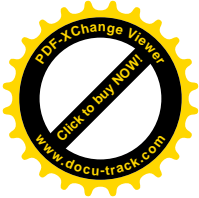
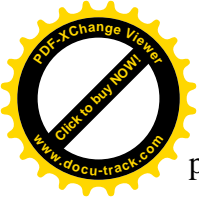
reinterpretation did not take place for the [a] lengthening across the EINE_{a+} series. Although the percentages of not-matching judgements also increased slightly but significantly when the [a] was lengthened by more than 80ms (in stimuli 5-7), the judgements of the the EINE_{a+} stimuli remained predominantly matching, i.e. *eine rote*. Compared with this, [ɪn^j] lengthening of just 20ms in the EINE_{in+} series was already sufficient to cause a significant increase in *eigentlich* identifications, and extending the [ɪn^j] section by another 40ms triggered almost exclusively *eigentlich* identifications.

These conclusions match well with the empirical corpus data. The natural productions of highly reduced *eigentlich* in the Kiel Corpus of Spontaneous Speech showed a lengthening of the palatal gliding section of more than 60ms compared with considerably less reduced *eigentlich* productions (cf. 1.3). [ɪn^j] lengthening of the same magnitude was required in Experiment 2 to create an unequivocal *eigentlich* cue in the EINE_{in+} stimuli. That the *eigentlich* identification was already raised by a 20ms [ɪn^j] lengthening shows that listeners can be very sensitive to duration changes in speech. Studies estimating the Just Noticeable Difference (JND) for variation of consonant and vowel durations in utterances also found that changes of 20ms are detectable (cf. Klatt and Cooper 1975; Bochner et al. 1988). Finally, it is worth noting that 60-80ms roughly corresponds to the average durations of sound segments in spontaneous speech, cf. Crystal and House (1988), Simpson (1998). Thus, local lengthening of this order of magnitude is likely to have an effect on the number of perceived sound portions. This could explain why strong lengthening increased the not-matching judgements even if it concerned the [a] section. Of course, unlike in the EINE_{in+} series, it is possible that the increase of not-matching judgements in stimuli 4-5 of the EINE_{a+} series is not an indicator of *eigentlich* identifications. However, the direct identification findings of Experiment 1 are in favour of this possibility (cf. Tab.1). Therefore, follow-up studies must examine the perceptual effects of strong local lengthening and their interplay with other components of the phonetic essence in the identification of highly reduced words in more detail.

2.5 Experiment 3: indirect identification test with formal question context

2.5.1 Preliminaries

Experiments 1 and 2 have shown that articulatory prosodies that represent residuals of highly reduced words can trigger the identification of a word, even if there is no semantic bias towards this word in the utterance context. Hence, our findings differ from the conclusion of Ernestus et al. (2002), according to which such semantic support of existing acoustic cues is essential in the



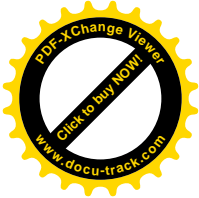
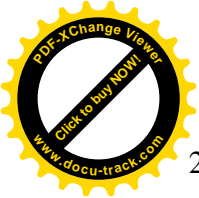
1 perception of highly reduced words. However, this need not mean that there are no context
2 effects at all. For example, it is possible that the wording of the stimuli is affected by different
3 speaking style contexts. A speaking style is characterized by a particular global reduction level.
4 In the context of such a global reduction level an expectation may be created concerning the
5 reduction level in the stimuli, which may in turn modulate listeners' perception of the articulatory
6 prosodies in the stimuli. This idea is tested in Experiment 3.
7
8
9

10 2.5.2 Procedure and participants

11 The third experiment represents a repetition of the second experiment with one modification. The
12 casual, strongly reduced context precursor *wieviele willst Du?* 'how many do you want?' was
13 replaced by a less reduced and hence more formal, carefully pronounced variant of this question.
14 If this new question precursor creates an expectation-based context effect in the stimuli, it will
15 bias the perception towards the less reduced variant *eine rote*. In terms of the indirect
16 identification task, the more formal, carefully pronounced question precursor will consistently
17 reduce the number of not-matching judgements, particularly for the EINE_{in+} stimuli.
18
19
20
21
22
23
24
25
26

27 The new variant of *wieviele willst Du?* was produced by the same female Standard German
28 speaker as in Experiment 2. The intonation contour remained constant in the sense that it again
29 consisted of a rising prenuclear pitch accent across the initial syllable *wie-*, followed by a fall and
30 a high-ending nuclear pitch-accent rise that starts around the onset of the vowel [ɪ] in *willst*. The
31 f₀-ranges and alignments of the pre-nuclear and nuclear rises differed less than 20Hz and 20ms
32 between the casual and formal variants of *wieviele willst Du?*. However, as a result of its more
33 careful pronunciation the formal variant is overall 170ms longer than the casual variant.
34
35
36
37
38
39
40
41
42

43 The concatenation of the constant question context with the 14 individual stimuli, the repetition
44 and randomization of the question - stimulus pairs, their presentation and judgement procedures,
45 as well as the instruction of the listeners were identical with Experiment 2. However, in order to
46 exclude potential learning effects or other response biases introduced by Experiment 2,
47 Experiment 3 was run with a different group of 20 listeners who were naïve concerning the aim
48 and background of the experiment. The 13 female and 7 male subjects (average age 25.1 years)
49 were again recruited from the undergraduate students of the Department of General and
50 Comparative Linguistics at the University of Kiel. None of the subjects participated in
51 Experiments 1 or 2; all of them were native speakers of German and reported normal hearing.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

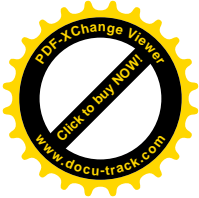
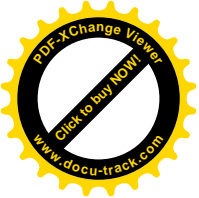


2.5.3 Results and discussion

The results of Experiment 3 were combined with those of Experiment 2 for the statistical analysis by means of a repeated-measures ANOVA in which the two within-subject factors ‘degree of lengthening’ and ‘type of lengthened section’ were complemented by the between-subject factor ‘context pronunciation’ with its two levels informal vs. formal. The three-factor ANOVA showed highly significant main effects for the degree of lengthening ($F_{(6,228)}=301.449$; $p<0.001$; $\eta_p^2=0.888$) and the type of lengthened section ($F_{(1,38)}=991.208$; $p<0.001$; $\eta_p^2=0.963$). The context pronunciation also had a highly significant main effect on the not-matching judgements ($F_{(1,38)}=86.552$; $p<0.001$; $\eta_p^2=0.695$). Moreover, all pairwise interactions between the three factors were significant (degree * type of lengthening: $F_{(6,228)}=104.167$; $p<0.001$; $\eta_p^2=0.733$; degree of lengthening * context $F_{(6,228)}=6.014$; $p<0.001$; $\eta_p^2=0.137$; type of lengthened section * context: $F_{(1,38)}=91.530$; $p<0.001$; $\eta_p^2=0.707$). The same was true for the three-way interaction ($F_{(6,228)}=12.959$; $p<0.001$; $\eta_p^2=0.254$). Even though the effect size of the three-way interaction was relatively small, we split up the three-factor ANOVA along the type of lengthened section into a pair of two-factor repeated-measures ANOVAs that tested the effects of degree of lengthening and context pronunciation separately for the EINE_{in+} and EINE_{a+} series. Our prediction (cf. 2.5.2) is that the formal question context of Experiment 3 counteracts the lengthening in that it reduces the number of not-matching judgements; therefore we added a priori repeated contrasts for the two-factor interactions, which we expected to become significant.

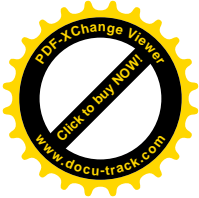
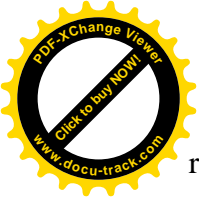
//////////INSERT FIGURE 5 ABOUT HERE//////////

As can be seen in Figure 5, the overall judgement pattern created by the 14 question - stimulus pairs and the 20 listeners of Experiment 3 is similar to the pattern of Experiment 2. The successive lengthening of the palatal [ɪnʲ] section across stimuli 1-7 of the EINE_{in+} series caused a clear change in the majority of judgements from matching to not matching. Such a change did not occur across the EINE_{a+} series, in which the [a] section was successively lengthened. As in Experiment 2, the proportion of not-matching judgements for EINE_{a+} stimuli increased as well across the series, but did not substantially exceed the 30% level. Correspondingly, we found a highly significant main effect of the within-subject factor degree of lengthening in the ANOVAs of both the EINE_{in+} series ($F_{(6,228)}=320.709$; $p<0.001$; $\eta_p^2=0.894$) and the EINE_{a+} series ($F_{(6,228)}=36.718$; $p<0.001$; $\eta_p^2=0.491$). However, reflecting the judgement curves in Figure 5, the effect size of the lengthening was roughly twice as high in the EINE_{in+} series.



1 Furthermore, a comparison of Figure 5 with Figure 4 shows clearly that the stimuli in
2 Experiment 3 triggered fewer not-matching judgements, particularly in the case of the EINE_{in+}
3 series. The change from matching to not matching across the EINE_{in+} stimuli occurred after
4 stimulus 2 in Experiment 2, but after stimulus 4 in Experiment 3. Stimulus 7, which has the most
5 extensive lengthening of the palatal [ɪn^ɪ] section (120ms), yielded around 80% not-matching
6 judgements in Experiment 3. In Experiment 2, a similar not-matching proportion was already
7 reached by stimulus 3, in which the [ɪn^ɪ] section shows only a third of the most extensive
8 lengthening (40ms). These observations are reflected in the ANOVA of the EINE_{in+} series in a
9 highly significant interaction between the degree of lengthening and the pronunciation of the
10 question context ($F_{(6,228)}=16.988$; $p<0.001$; $\eta_p^2=0.551$). Moreover, the repeated contrasts that
11 compared each degree of lengthening with the previous one across the two context conditions
12 showed that the interaction mainly relates to the more or less rapidly increasing not-matching
13 levels in stimuli 1-2 ($F_{(1,38)}=14.572$; $p<0.001$; $\eta_p^2=0.277$), 2-3 ($F_{(1,38)}=4.234$; $p=0.047$;
14 $\eta_p^2=0.100$), and 6-7 ($F_{(1,38)}=6.480$; $p=0.015$; $\eta_p^2=0.146$). On the other hand, the negligible visual
15 differences that can be found for the EINE_{a+} series in Figures 4-5 are not reflected by
16 significances in the corresponding ANOVA either for the interaction between degree of
17 lengthening and context or for any of the repeated contrasts (the repeated contrast for stimuli 4-5
18 across the two context conditions approximated significance). Overall, the only significant effect
19 that showed up for the ANOVA on the EINE_{a+} series relates to the main effect of lengthening,
20 which is, in turn, detailed by the post hoc comparisons between its 7 levels. However, unlike in
21 Experiment 2, the only significant increase in not-matching judgements did not occur in stimuli
22 4-5, but between stimuli 5-6 ($p=0.031$; with Bonferroni correction, cf. 2.4.3).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43 As was argued in the discussion of Experiment 2, the findings of the direct identification in
44 Experiment 1 allow translating the matching/not-matching judgements into *eine rote* or
45 *eigentlich 'ne rote* perceptions. If the same is done for Experiment 3, then the more formal,
46 carefully pronounced question precursor in fact created a perceptual bias towards *eine rote*. In the
47 EINE_{a+} series this bias manifested itself in the delay of the significant increase of not matching
48 from stimuli 4-5 in Experiment 2 to 5-6 in Experiment 3. However, the bias especially concerned
49 the EINE_{in+} stimuli. Looking at the discrepancies in not-matching levels for the stimuli with the
50 same numbers in Experiments 2 and 3, the strongest bias is found for the stimuli in the centre of
51 the lengthening continuum (around stimulus 3), which contained more ambiguous cues to *eine*
52 *rote* or *eigentlich 'ne rote*. The perceptual bias is in line with an expectation-based effect of the
53 speaking-style context. That is, if an utterance has a low reduction level, then listeners are more
54
55
56
57
58
59
60
61
62
63
64
65

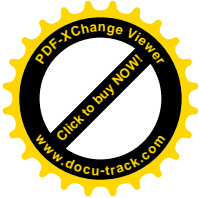
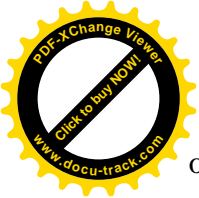


1 reluctant to perceive articulatory prosodies in the following utterance as being related to the
2 phonetic essence of a highly reduced word. Of course, this initial indirect evidence of a context
3 effect of speaking style on word identification must be further supported by direct evidence, for
4 example, by using the dictation-task approach of Experiment 1 or a variant of the imitation
5 procedure (cf. Pierrehumbert and Steele 1989; Niebuhr 2009). In any event, the existence of a
6 context effect of speaking style requires that listeners are able to perceive and to differentiate
7 degrees of reduction.
8
9

10
11
12 This key requirement is supported by the fact that variation in the degree of reduction is involved
13 in a number of communicative functions that are all in different ways interactional. For example,
14 Kirchner (1998) found that phonetic bracketing, i.e. the insertion of short complementary
15 utterances into the surrounding speech, is not only done by lowering and flattening the pitch
16 course, but also by using a higher degree of reduction relative to the surrounding speech. Plug
17 (2005) showed that utterances expressing disagreement with the dialogue partner also differ from
18 the speech-reduction level of the dialogue partner, typically in the direction of less reduction.
19 Local (2003) noted that the phrase *I think* is more strongly reduced when it is used in a de-
20 lexicalized way as a marker that hedges the meaning of the preceding utterance(s). According to
21 Hawkins (2003), the reduction of the English utterance *I do not know* to [ə̃ə̃ə̃] not only signals
22 that the speaker has no answer to the question of the dialogue partner. It can also convey that s/he
23 is unwilling to cooperate with the dialogue partner in finding an answer. Similarly, by reducing
24 the German greeting *guten morgen* to [g̃ũmõ], the speaker can express her/his unwillingness to
25 start a conversation. In view of such functional exploitations of reduction levels in speech
26 communication, reduction cannot be understood as simply striving for articulatory economy (cf.
27 Simpson 2001). However, segmental reduction is probably not the only means for conveying the
28 sketched functions. As implied by the example of phonetic bracketing, segmental reduction may
29 have repercussions on intonation, or there could be combinatory restrictions between the degree
30 of reduction and pitch patterns within the prosodic phrase. This is an interesting perspective for
31 new research at the segment-prosody intersection.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 **3. Conclusions and outlook**

53
54 The study of phonetic reduction of lexical items has traditionally been carried out in a segmental
55 phonemic frame of canonical forms. This approach mirrors the pivotal role attributed to the
56 phoneme in production and perception studies generally. Local phonetic detail, associated with
57 remaining segments in the reduced input form, may assist in this mediating process. But the
58 framework does not consider non-linear articulatory prosodies as phonetic components in their
59
60
61
62
63
64
65

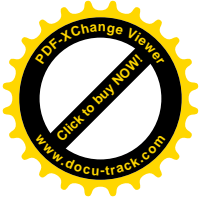
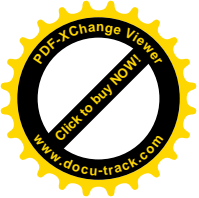


own right, with variable positions and extensions in a segmental chain, which capture the phonetic essence of the whole form class of a lexical item.

Such articulatory prosodies can be found, and systematically accounted for, in the phonetic manifestations of lexical forms in spontaneous speech. Since production features are mapped onto perception it is to be assumed that these articulatory prosodies play an important role in the identification of words as indices of their phonetic essence. This means that the traditional phoneme-based paradigm in phonetic production and perception analysis needs to be adjusted, and the category of articulatory prosody admitted as a variable of investigation.

This is the theoretical stance of this paper. The experiments were to shed new light on the perception of reduced speech with regard to articulatory prosodies. The major findings can be summarized as follows. Contrary to previous conclusions, it was demonstrated that the identification of highly reduced words which partly or entirely lack a separate segmental-phonetic representation need not be inferred from the syntactic/semantic context. The words can still be sufficiently coded and hence directly identified by means of articulatory prosodies that preserve major phonetic characteristics of the “missing” segments and hence retain the phonetic essence of that word. Our experiments dealt with two articulatory prosodies of the German word *eigentlich*, duration and palatality, which were found in previous acoustic analyses. We showed that they represent essential phonetic features for the perceptual identification of the word *eigentlich*, as they can trigger a reinterpretation of the stimulus utterance *eine rote* as *eigentlich 'ne rote*. Lengthening per se was not sufficient for this reinterpretation to occur. It had to be linked to the palatal glide section of the nasalized diphthong [aĩ].

Moreover, we provided indirect evidence for the existence of a context effect of the speaking style on word identification. Even though articulatory prosodies can be cues to the identification of highly reduced words, their perception is modulated by the reduction level of the surrounding utterance context. In our case this context was constituted by a preceding question that was produced either highly reduced or in a more careful fashion with a lower reduction level. Our experimental data suggest that the lower reduction level in the question context suppressed the perception of articulatory prosodies as indicators of highly reduced words. Such a speaking-style effect presupposes that listeners can differentiate different degrees of reduction, at least two very dissimilar ones. This is inconsistent with a strict understanding of phonemic restoration that makes “reduced forms difficult to distinguish from their non-reduced counterparts” (Kemps et al. 2004:117). That is, non-segmental phonetic detail of the word must be processed and stored.

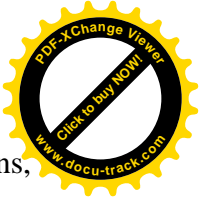
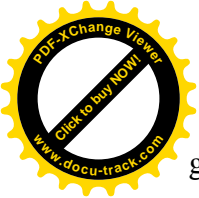


1 In addition to finding more direct evidence for an effect of speaking-style context on word
2 identification, follow-up studies need to explore the notion of phonetic essence further both in
3 production and perception. As regards the former, more detailed studies of articulatory prosodies
4 and their interplay are necessary. In the domain of perception, a follow-up experiment could test
5 whether, in view of the production data of German *eigentlich* (cf. 1.3), identification can also be
6 triggered by increasing the second-formant frequency of the gliding section in the utterance-
7 initial diphthong [aĩ]. Furthermore, it will be interesting to carry out a perceptual investigation
8 into the phonetic essence in reduced forms of the Dutch equivalent of *eigentlich*, i.e. *eigenlijk*.
9 Compared with German *eigentlich* [aɪŋtliç], Dutch *eigenlijk* [ɛiχələk^h] is characterized by
10 lower and more back, i.e. non-palatal articulations, and it has no nasal element. This is mirrored
11 in highly reduced forms like [ɛɛŋ^ɾ] vs. German [aĩ]. Thus, if the concept of phonetic essence
12 holds, different articulatory prosodies are to be expected to guide lexical identification of
13 extremely reduced forms of *eigenlijk* by Dutch listeners. In the case of the *Ihnen* example in 1.1,
14 an initial perception experiment has just shown that the identification of this word can be
15 triggered just by adding palatality, i.e. without a coinciding increase in duration. So, duration is
16 involved in the differences between the phonetic essence of *eigentlich* and the one of *Ihnen*. This
17 fact lends further support to our characterization of duration as a separate articulatory prosody.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33
34 In general, the investigation of spontaneous speech should be done under the premise that –
35 contrary to its semantic implication – reduction cannot be simply conceptualized as a lowering of
36 the information value of the speech code. This reduction concept rests on the idea that speech
37 communication is based on segmental-phonemic units. Realizations of words in connected
38 speech are then the result of assimilation and elision processes that delete phonological features
39 or segments and thus reduce the cues to word identification. Such a segmental-phonemic
40 perspective misses the information value of phonetic detail of articulatory prosodies in the
41 production and perception of words. So, the ultimate message of this paper is that phonetic
42 research would benefit greatly from a change of its phoneme-based paradigm, leading to a new
43 type of sound abstraction in the modelling of speech that is not limited to segmental units but
44 also includes long articulatory components, over and above pitch, loudness and rate prosodies, as
45 basic elements for production and perception.
46
47
48
49
50
51
52
53
54
55
56
57

58 **4. Acknowledgements**

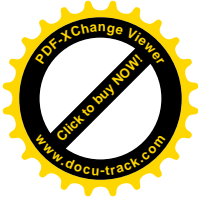
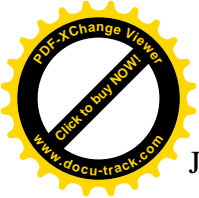
59 First of all, we would like to thank our reviewers. Their detailed and stimulating comments and
60 suggestions contributed to improving the argumentation of the paper. Moreover, we are very
61
62
63
64
65



grateful to Meghan Clayards, Noel Nguyen, Christine Meunier, Gareth Gaskell, Sarah Hawkins, Annett Jorschick, Ernst Dombrowski, and many people of the research-training network “Sound to Sense” for vivid and inspiring discussions about speech reduction and statistics.

5. References

- Bochner, J.H., Snell, K.B., & MacKenzie, D.J. (1988). Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. *JASA* 84, 493-500.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5, 341-345.
- Crystal, D.H. & House, A.S. (1988). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America* 83, 1574-1585.
- Ernestus, M. (2000). *Voice assimilation and segmental reduction in Dutch*. PhD dissertation, University of Utrecht, The Netherlands.
- Ernestus, M., Baayen, H.R., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language* 81, 162-173.
- Firth, J.R. (1948). Sounds and prosodies. *Transactions of the Philological Society* 1948, 127-152.
- Gadet, F. (1992). *Le français populaire*. Paris : PUF.
- Gaskell, G. & Marslen-Wilson, W.D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 144-158.
- Gaskell, G. & Marslen-Wilson, W.D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language* 44, 325–349.
- Gimson, A.C. (1962). *An introduction to the pronunciation of English*. London: Edward Arnold (8th ed. revised by A. Cruttenden (2008) *Gimson’s pronunciation of English*. London: Hodder Education/Hachette Livre.)
- Gow, D.W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception & Performance* 28, 163-179.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31, 373-405.
- Heid, S., & Hawkins, S. (2000). An acoustical study of long-domain /r/ and /l/ coarticulation. *Proc. 5th seminar on speech production: models and data, Kloster Seeon, Germany*, 77–80.
- Holst, T., & Nolan, F. (1996). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (Eds.) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence* (pp. 315-333). Cambridge: CUP
- IPDS (1995). *The Kiel Corpus of Spontaneous Speech, volume 1*, CD-ROM#2. Kiel: IPDS.
- IPDS (1996). *The Kiel Corpus of Spontaneous Speech, volume 2*, CD-ROM#3. Kiel: IPDS.



1 Jones, D. (1997). *English Pronouncing Dictionary*. 15th ed. (P. Roach, J. Hartman, eds.).
2 Cambridge: Cambridge University Press.

3 Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H.R. (2004). Processing reduced word forms:
4 The suffix restoration effect. *Brain and Language* 90, 117-127.

5 Kirchner, R. (1998). *An effort-based approach to consonant lenition*. Ph.D. dissertation,
6 University of California at Los Angeles.

7 Klatt, D.H. & Cooper, W.E. (1975). Perception of segment duration in sentence contexts. In A.
8 Cohen & S.G. Nootboom (Eds.) *Structure and Process in Speech Perception* (pp. 69-89).
9 Springer: New York.

10 Kleber, F. (2006). Form and function of falling pitch contours in English. *Proc. 3rd International*
11 *Conference of Speech Prosody, Dresden, Germany*, 61-64.

12 Kohler, K.J. (1990). Segmental reduction in connected speech in German: phonological facts and
13 phonetic explanations. In W. Hardcastle & A. Marchal (Eds.) *Speech production and speech*
14 *modelling* (pp. 69-92). Dordrecht: Kluwer.

15 Kohler, K.J. (1994). Complementary phonology: a theoretical frame for labelling an acoustic
16 data base of dialogues. *Proc. ICSLP94*, vol. 1, 427-430, Yokohama.

17 Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag
18 (2nd ed.)

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

21 Kohler, K.J. (1999). Articulatory prosodies in German reduced speech. In Proc. XIVth ICPhS,
22 volume 1, 89-92, San Francisco.

23 Kohler, K.J. (2001). Articulatory dynamics of vowels and consonants in speech communication.
24 *Journal of the International Phonetic Association* 31, 1-16.

25 Kohler, K., Pätzold, M., & Simpson, A. (1995). *From scenario to segment. The controlled*
26 *elicitation, transcription, segmentation and labelling of spontaneous speech*. IPDS: Kiel.

27 Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

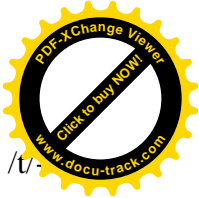
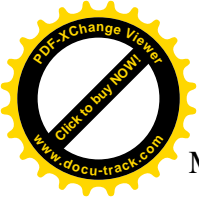
28 Landau, S. & Everitt, B. (2004). *A handbook of statistical analysis using SPSS*. Boca Raton:
29 Chapman & Hall.

30 Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

31 Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents.
32 *Journal of Phonetics* 31, 321-339.

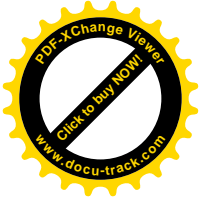
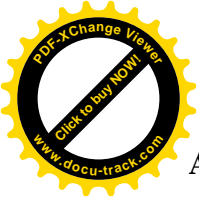
33 Manuel, S.Y. (1992). Recovery of “deleted” schwa. *Perilus* 14, 115-118.

34 Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception:
35 Evidence for early compensation. *Perception and Psychophysics* 65, 956-969.



- 1 Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/
2 elision in Dutch. *Journal of Phonetics* 34, 73-103.
- 3 Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there:
4 Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of*
5 *Memory and Language* 59, 133-152.
- 6 Nash, R. & Mulac, A. (1980). The intonation of verifiability. In L.R. Waugh & C.H. van
7 Schooneveld (Eds.) *The melody of language: Intonation and prosody* (pp.219 –241). Baltimore:
8 University Park Press.
- 9 Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In D.R. Ladd &
10 G.J. Docherty (Eds.) *Papers in Laboratory Phonology 2* (pp. 261–280). Cambridge: CUP.
- 11 Niebuhr, O. (2007). The signalling of German rising-falling intonation categories – the interplay
12 of synchronization, shape, and height. *Phonetica* 64, 174-193.
- 13 Niebuhr, O. (2009). f0-based rhythm effects on the perception of local syllable prominence.
14 *Phonetica* 66, 95-113.
- 15 Niebuhr, O., Meunier, Ch., & Lancia, L. (2009). The role of the vowel context in the
16 differentiation of French /sS/ and /Ss/ sequences. Talk given at the PaPI 2009 conference, Las
17 Palmas, Spain. <http://www.ipds.uni-kiel.de/on/>
- 18 Pierrehumbert, J.B. & Steele, S.A. (1989). Categories of tonal alignment in English. *Phonetica*
19 46, 181-196.
- 20 Plug, L. (2005). From words to actions: The phonetics of eigenlijk in two communicative
21 contexts. *Phonetica* 62, 131-145.
- 22 Simpson, A.P. (1998). *Phonetische Datenbanken des Deutschen in der empirischen*
23 *Sprachforschung und der phonologischen Theoriebildung*. IPDS: Kiel.
- 24 Simpson, A.P. (2001). Does articulatory reduction miss more patterns than it accounts for?
25 *Journal of the International Phonetic Association* 31, 29-39.
- 26 Snoeren, N.D., Hallé, P.A., & Segui, J. (2006). A voice for the voiceless: Production and
27 perception of assimilated speech in French. *Journal of Phonetics* 34, 241-268.
- 28 Weilhammer, K. & Rabold, S. (2003). Durational aspects in turn taking. *Proc. 15th International*
29 *Congress of Phonetic Sciences, Barcelona, Spain*, 931-934.
- 30 Wells, J.C. (1990). *Pronunciation Dictionary*. London: Longman.
- 31 WÖRTERBUCH DER DEUTSCHEN AUSSPRACHE (1969). H. Krech et al. (eds.), München:
32 Max Hueber.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



Appendix

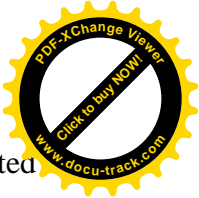
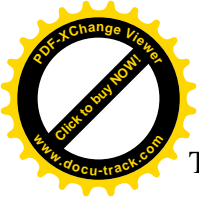
The three experimental texts in Experiment1. The stimulus slots underlined.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

(A: Poker) Stefan und Ernst sitzen schon seit Stunden in einem dunklen, rauchigen Kellerraum beim Poker. Nach einiger Zeit wird es Peter zu bunt. “Das ist unfair. Du hast bis jetzt jedes Pokerspiel gewonnen, weil Du immer die richtigen Karten ziehst. Und? Was brauchst Du denn diesmal?” “STIMULUS. Aber ich fürchte, von dieser Farbe sind nicht mehr viele Karten im Spiel. Sieht so aus, als würde ich diese Runde verlieren.“

(B: Election) Nach dem erwarteten Ausgang einer Abstimmung im Bundestag unterhalten sich zwei befreundete Abgeordnete. “Unglaublich. Das Gesetz hat tatsächlich eine Mehrheit bekommen. Hätte nie gedacht, dass sie auch dafürstimmt.“ “Tja, Du kennst sie ja. Ist doch nicht das erste Mal, dass sie so einen Alleingang hinlegt. STIMULUS. Und doch so viel Wert wie tausend Schwarz-Gelbe.“

(C: Wine) Stefan und Ernst genießen ihre erste Weinprobe an der Mosel. Immer abwechselnd trinken sie einen Rot- und Weißwein nach dem anderen. “Weinproben sind einfach etwas Tolles. Man kann soviel durchprobieren. Welche Weinflasche ist als nächstes dran?“ “STIMULUS. Aber wenn ich an den tollen Weißwein von eben denke, dann lass uns lieber noch einmal so eine Flasche aufmachen.“



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1: Absolute frequencies and percentages (in brackets) of *eigentlich* transliterations elicited by the three stimuli within the experimental texts ‘poker’, ‘election’, and ‘wine’ that correspond to the three experimental sessions with subgroups of 15 listeners (i.e. n=15 for each cell). The grey column presents the total frequencies and percentages that resulted across all three experimental texts/sessions (i.e. n=45 for each cell).

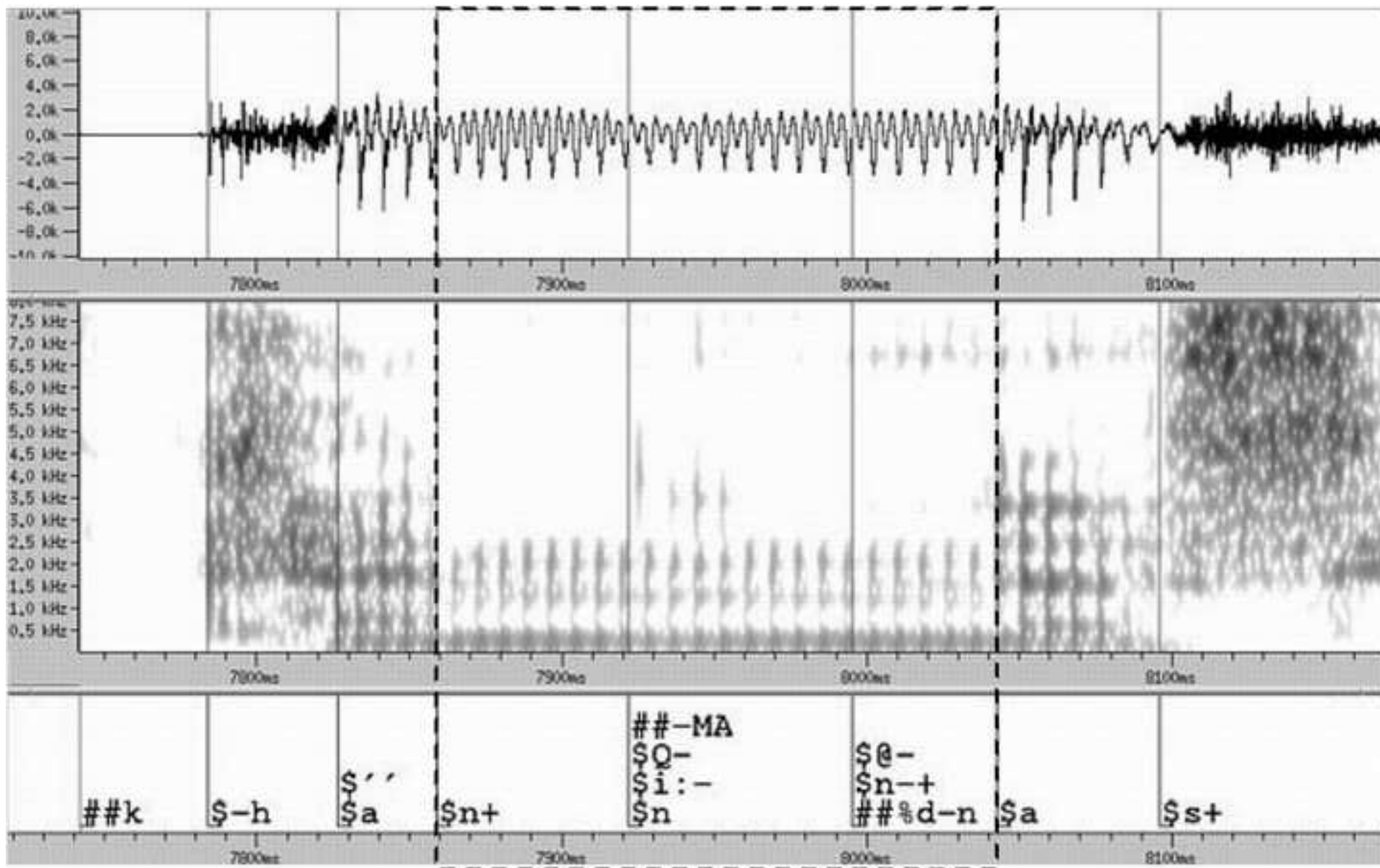
Figure 1. Speech wave, spectrogram, and aligned SAMPA labels in the *Kiel Corpus* annotation system, for the spontaneous German utterance *ich kann Ihnen das ja mal sagen* ‘I can mention this to you’. The dotted box rims the long alveolar nasal (ca. 180ms) in between [ka__as] .

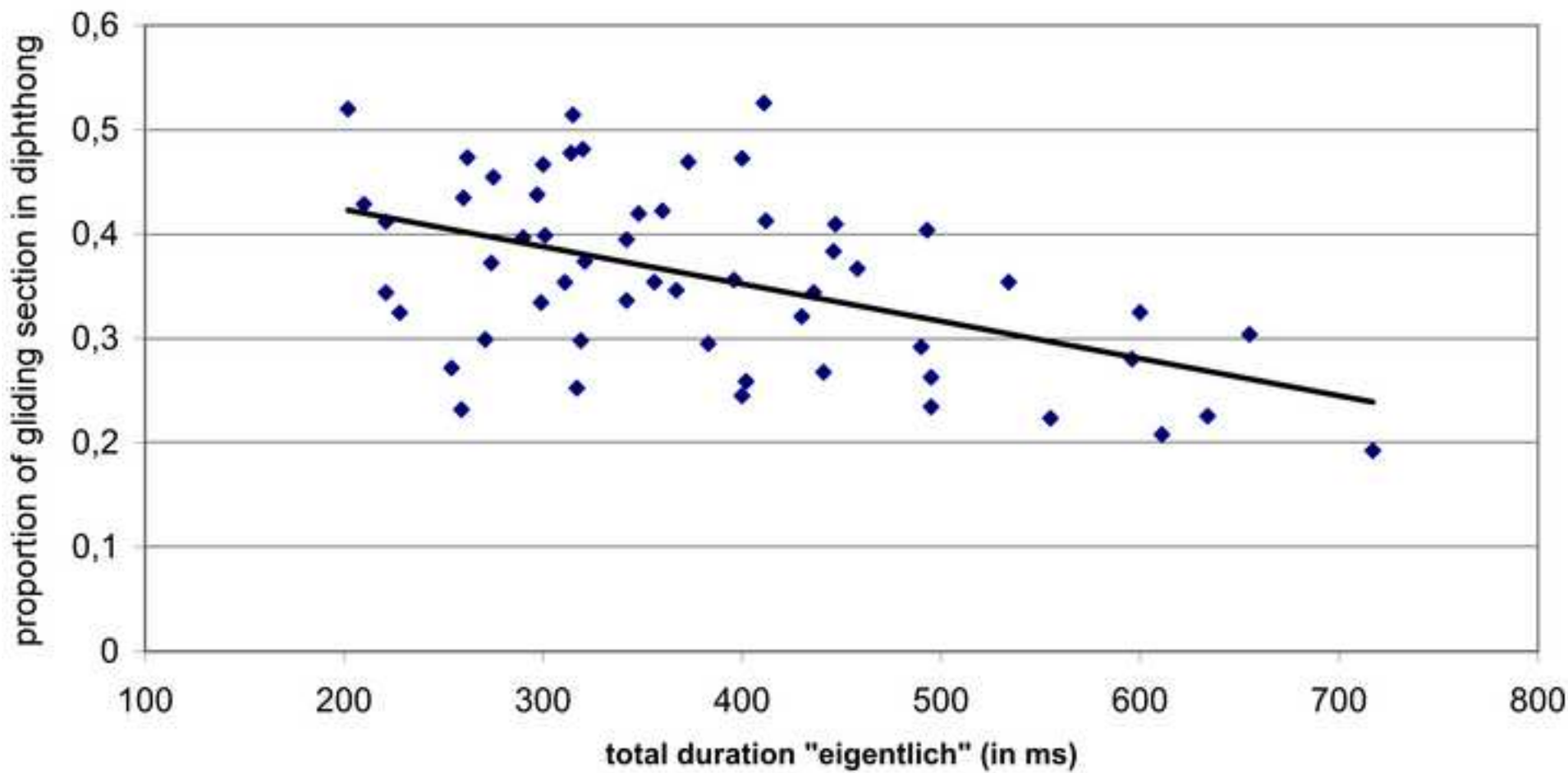
Figure 2. Correlation between the overall word duration and the proportion of the gliding section in the diphthong of *eigentlich* in the *Kiel Corpus* of Spontaneous Speech; n=56.

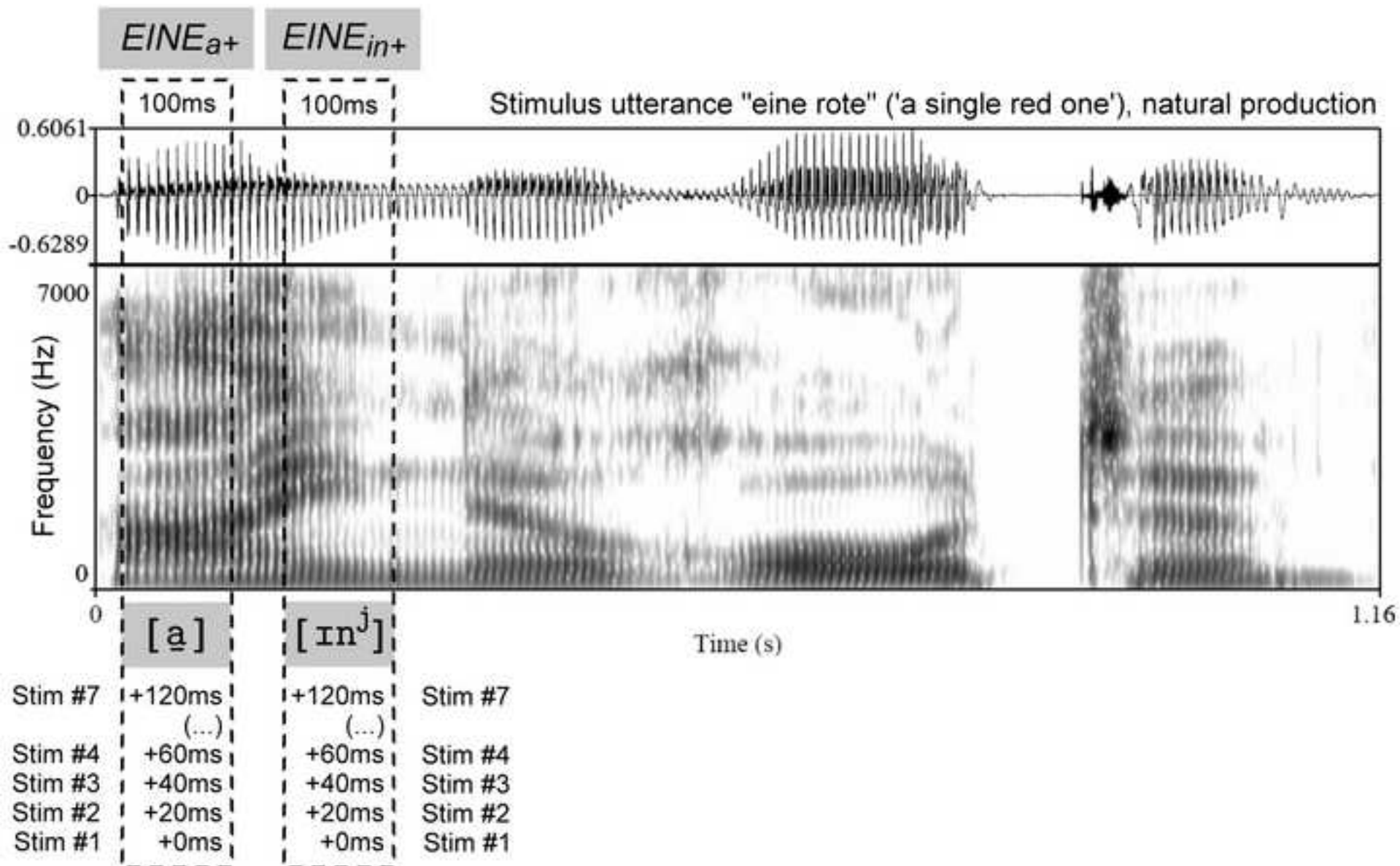
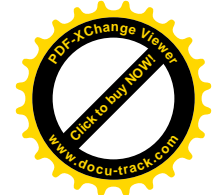
Figure 3. Generation of the two stimulus series $EINE_{in+}$ and $EINE_{a+}$ by stepwise linear lengthening the [ɪn^j] or [a] sound sections in the initial *eine* of the naturally produced utterance *eine rote*. Oscillogram (top) and spectrogram (0.5-7kHz, bottom) show the original temporal structure.

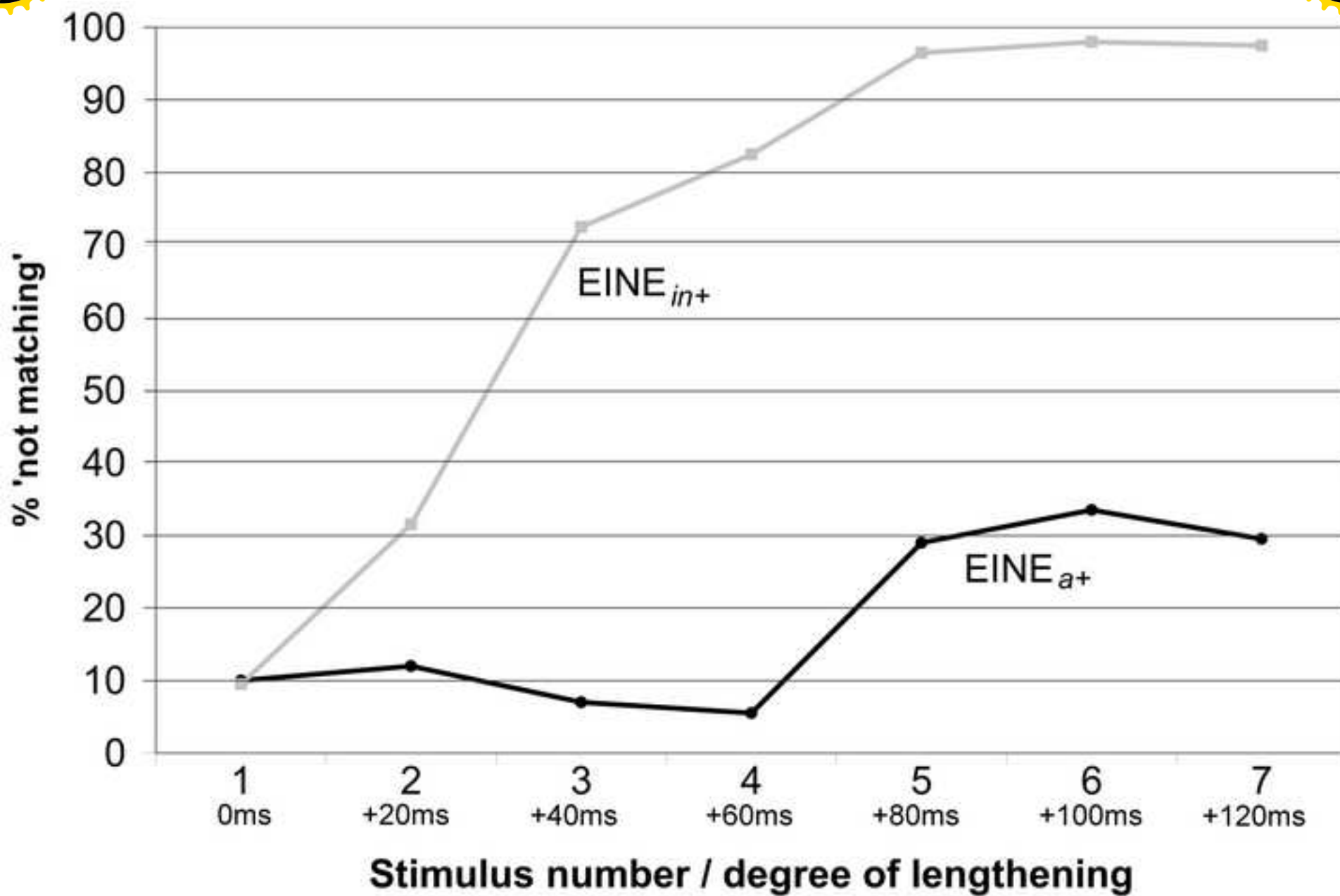
Figure 4. Percentages of stimuli judged as not matching in the context of the casual question precursor *wieviele willst Du?* ‘how many do you want?’. Stimuli of the $EINE_{in+}$ and $EINE_{a+}$ series are represented by the grey or black curves, respectively. Each data point represents 10 (repetitions) x 20 (listeners) = 200 judgements.

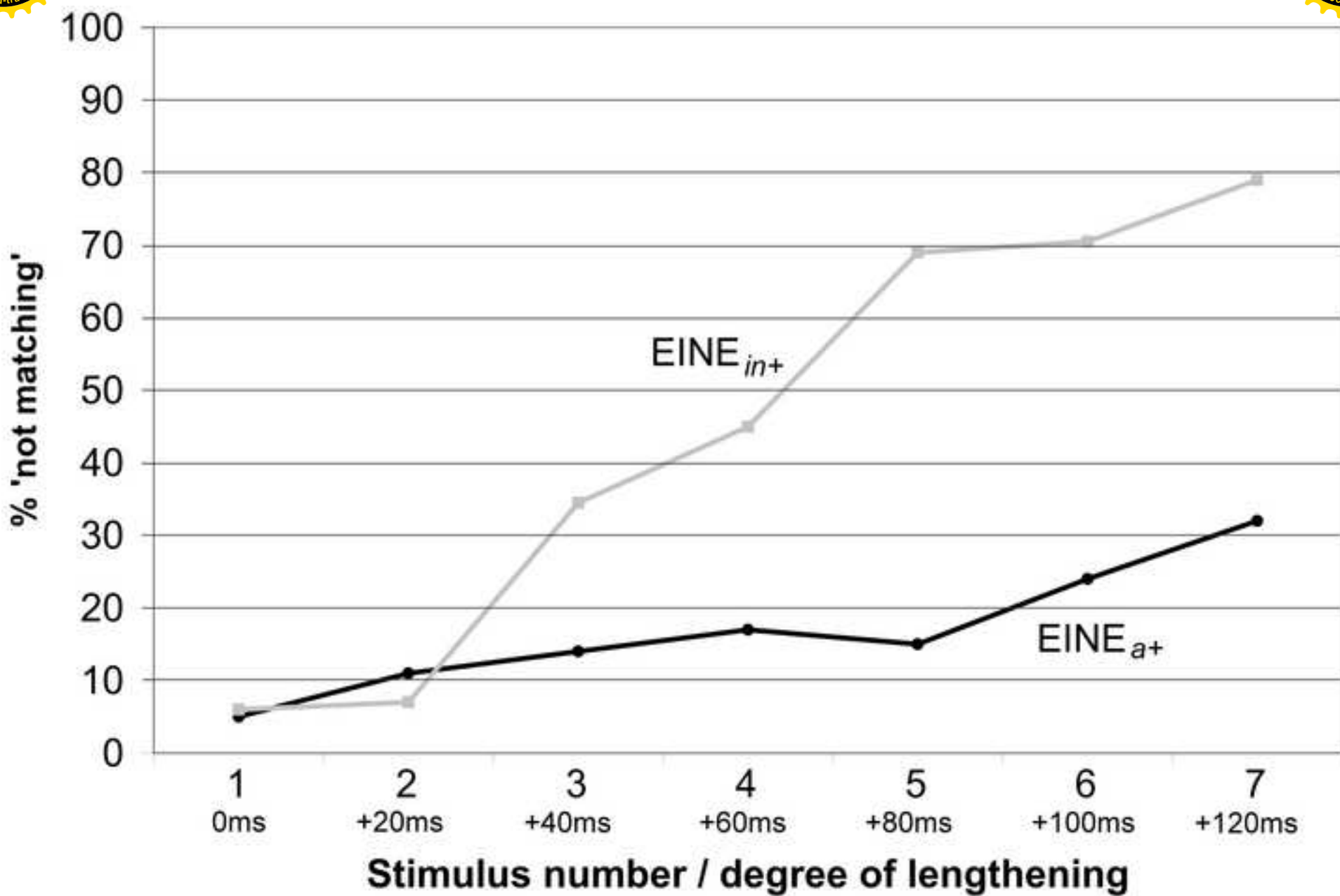
Figure 5. Percentages of stimuli judged as not matching in the context of the formal, carefully pronounced question precursor *wieviele willst Du?* ‘how many do you want?’. Stimuli of the $EINE_{in+}$ and $EINE_{a+}$ series are represented by the grey or black curves, respectively. Each data point represents 10 (repetitions) x 20 (listeners) = 200 judgements.











"eigentlich"
 occurrences

Experimental texts

Σ (%)	poker	election	wine	total Σ (%)
Stim. 1	1 (6.7)	2 (6.7)	2 (13.3)	5 (11.1)
Stim. 7 EINE _{in+}	14 (93.3)	15 (100)	15 (100)	44 (97.8)
Stim. 7 EINE _{a+}	3 (20)	3 (20)	5 (33.3)	11 (24.4)