

## On the Domain of Auditory Restoration in Speech

Oliver Niebuhr (oliver.niebuhr@linguistik.uni-kiel.de)

Department General & Comparative Linguistics, Leibnizstrasse 10

University of Kiel, 24098, Germany

### Abstract

The paper presents two perception experiments that start from the well-established phenomenon of auditory restoration and address the occurrence and nature of this phenomenon in German reduced speech. The experimental stimuli based on sound patterns that can be decoded as either disyllabic nouns or highly reduced strings of function words whose canonical reference forms differ in the number of syllables and/or phonemes. Comparisons of these stimuli with [ə]-like hum sounds of different durations in AX tests suggest that reduced speech is perceptually restored to fuller (though maybe not canonical) forms, however, rather on a syllable than on a phoneme basis. In this way, the results raise further doubts about the central role of the phoneme in spoken communication.

**Keywords:** phonemic restoration; speech perception; speech reduction, German, phonetic detail; phoneme; syllable.

### Introduction

Auditory restoration of missing sound sections in the speech signal is a well-known, robust phenomenon. In his seminal experiment Warren (1970) removed speech sounds from utterances and replaced them by noise (e.g., a cough). He found that listeners perceived both the removed speech sound and the noise (which was, however, temporally not exactly locatable) in the stimuli. Warren's effect has been replicated and refined many times since then, showing that, for example, acoustic similarity between original and replaced sound, presence and type of semantic context, temporal coherence and spatial localization of noise and interrupted signal, multi-modal perception, or the ability of the noise to be a potential masker of the missing sound play a crucial role in the occurrence and the output of auditory restoration, cf. Warren & Sherman (1974), Bashford & Warren (1987), Samuel (1981), Samuel (1996), Shahin & Miller (2009) and Shinn-Cunningham & Wang (2008). Since virtually all experiments were about the removal/masking and restoration of single speech segments, the effect is typically referred to as phonemic restoration.

Recent psycholinguistic research (e.g., Kemps et al., 2004; Connine et al., 2008; Mitterer & Ernestus, 2006) applied the notion of phonemic restoration to reduced speech, motivated by the concept that speech production and perception relate to a mental lexicon in which words are represented as canonical/unreduced forms with a linear, phonemic make-up. Thus in speech production the canonical forms are reduced by means of assimilation and elision processes that result in "missing" phonemes; and in speech perception the reference to canonical forms causes the restoration of the "missing" phonemes (either before or even after word identification). However, in-depth investigations into the phonetic details of reduced speech accumulated

evidence that the essential phonetic characteristics of "missing" phonemes or even entire (function) words are still in the signal, but in the form of articulatory prosodies that are no longer tied to specific segmental units, cf. Kohler (1999), Niebuhr & Kohler (2011); Kohler & Niebuhr (2011). Moreover, it was found that also complete assimilations leave traces of the original sound features that are outside the assimilated sound itself, for example, in the preceding vowel, cf. Niebuhr & Meunier (2011). On the side of perception listeners are able to make use of all these non-linear and suprasegmental phonetic details in word identification (cf. also Hawkins 1995).

Compared with the stimulus conditions in the line of research initiated by Warren, (highly) reduced speech represents a completely different starting point for perception and cognitive processing. First, there is no need to reconstruct something, since the seemingly "missing" information can still be there; it just eludes a segmental perspective. Second, there are neither real nor potential interrupters or maskers that could motivate a restoration. Moreover, the tasks and/or the stimulus conditions of traditional phoneme restoration experiments force the listener to look from a segmental perspective onto the presented speech, which is not the case in natural everyday communication. For instance, the experiments of Kemps et al. (2004), Connine et al. (2008) and Mitterer & Ernestus (2006) that found evidence for phonemic restoration in reduced speech used phoneme-monitoring tasks and/or stimulated speech-orthography associations that are at least suitable to create perceptual biases towards segments and particular canonical reference forms. In line with this methodological deficiency, the results of Connine et al. (2008) suggest that perceptually restored morphs have no phonemically homogeneous reference forms.

In front of the sketched empirical background the present paper represents a first approach to revisit two fundamental questions: (1) Is auditory restoration really a "natural, highly frequent" process in the perception of reduced speech, as was claimed by Kemps et al. (2004:125)? The robust discrepancy between written and spoken language as well as the stunned reactions of people that are confronted with their own speech reductions are actually in favour of restoration processes. However, (2) if there is restoration, what is the point in basing it on phonemes that are (a) per definition meaningless units and that are (b) so loosely related to the actual non-linear, suprasegmental coding of reduced speech? More psychoacoustically-oriented experiments demonstrated that auditory restoration also occurs for sine waves interrupted by noise. So neither phonemes nor phonemic reference forms are prerequisites for restoration effects. For example, Ohala (1992) and others (Connine et

al., 2008) pointed out that larger sound sections like syllables would be much more effective basic units in the processing and perception of speech. Supporting this idea, the presented series of perception experiments on German, which avoid explicit reference to phonemic sound segments, yielded initial indirect evidence that auditory restoration does occur for highly reduced speech; however, more on a global syllabic than on a local phonemic basis.

## Method

### Disyllabic stimulus bases

Inspired by the method of Niebuhr & Kohler (2011) the stimuli of the present experiments based on phonetic sound patterns of German that can be readily identified as single lexical nouns or verb phrases. However, when followed by a pitch-accented content word, phonetic details that are contained in these sound patterns trigger immediate reinterpretations of the patterns as strings of highly reduced function words. The experiments were built on 6 of these potentially ambiguous patterns that reflect actual realizations found in corpora of spontaneous speech. Each pattern represents a disyllable with a CV<sub>s</sub>C.CV(C) structure and primary stress on the phonologically short vowel (V<sub>s</sub>) of the initial syllable: (1) [ˈnɔ̃ʔ.mɔ̃] *Norma*, a popular German discount chain; (2) [ˈkʰɔ̃ʔ.m<sup>wɪwʰ</sup>] *Kermit*, famous frog puppet from The Muppet Show; (3) [ˈvɪs̩.sɔ̃n] *Wissen* ‘knowledge’; (4) [ˈvɪs̩.mə] *Wismar*, a German hanseatic city; (5) [ˈhäm.m<sup>a</sup>] *Hammer*, ‘hammer’, (6) [ˈnɪm̩.zi] *Nimm sie*, ‘take it’.

### The SYLLABLE series

The first stimulus series (SYLLABLE series) used disyllables (1)-(3). Each of them was produced several times by a trained phonetician, aiming at a constant speaking rate, voice quality and F0 register. The recordings were made digitally (96 kHz, 32 bit, mono) in a soundproofed booth at the University of Kiel’s Institute of Phonetics and Digital Speech Processing (IPDS). The three disyllables that were selected as base stimuli evoked clear noun identifications. However, at the same time they were deliberately realized with phonetic details (e.g., subtle (sub-)segmental lengthening, suprasegmental secondary articulations like nasalization, rounding, palatalization, etc., cf. transcriptions above) that allowed reinterpretations of the disyllables in the presence of an appropriate trigger. Furthermore, the three selected disyllables showed similar overall prosodies, including almost identical overall durations that were facilitated by the similar segmental make-up CV<sub>s</sub>C.CV(C). The remaining marginal duration differences between the three disyllables were equalized to 650ms by means of *psola* resyntheses in *praat* (Boersma, 2001), which required linear overall lengthening or shortening of less than 15%.

In addition to the target disyllables, the speaker also produced the unequivocal disyllabic verb *gucken* ([ˈkʰw.ʊ.kʰ<sup>n</sup>.w]), ‘watch/look’, 700ms) with a clear rising pitch accent (H\*)

on the initial and a terminal falling pitch movement (L-%) on the final syllable. It was produced with the intention of forming prosodically coherent, naturally sounding utterances when attached to each of the target disyllables (1)-(3).

As is already implied by this procedure, two stimulus conditions were created on the basis of the selected disyllables. In the condition ISO, there were three stimuli with identical overall durations (650ms), represented by the isolated disyllables (1)-(3). In the condition WORD+, the constant *gucken* was attached to each disyllable, which also resulted in three stimuli with identical overall durations (1.350ms). However, different from the ISO condition, the verb *gucken* in the WORD+ condition is a semantic and prosodic trigger for the perceptual reinterpretations of the preceding disyllables as (1) *Nun wollen wir mal + gucken* ‘now let us see’, (2) *Können wir mit + gucken* ‘may we watch with you’, (3) *Willst du den + gucken* ‘do you want to watch it’. The crucial point is that the reinterpretations lead to utterances that show (1) 5 syllables, (2) 4 syllables, and (3) 3 syllables in their unreduced forms. If the fact that the reinterpreted stimuli are based on different syllable numbers is paralleled by a perceptual restoration effect, then the three stimuli of the WORD+ condition should be perceived with different lengths that increase from (3) over (2) to (1), despite their physically identical durations. In contrast, the control stimuli (1)-(3) of the ISO condition should not differ in perceptual length.

This was tested in two separate AX tests in which the X elements were represented by the three stimuli of either the ISO or the WORD+ condition. In both tests the A elements were phonetically constant [ə]-like ‘hum’ sounds that were generated by means of *praat* with a fundamental frequency (F0) of 100Hz, and with frequencies of 500Hz, 1500Hz, and 2500Hz for the first three formants (F1-F3). Eleven A elements with increasing overall durations were created for each of the two test conditions ISO and WORD+. The 11 durations differed in equal-sized steps of 75ms. Two A elements were shorter than the X stimuli (-75ms and -150ms), the third A element had the same duration as X, and 8 A elements were longer than X (+75ms to +600ms). Each X was combined with all 11 As. The stimulus pairs were introduced by a beep and followed by a pause of 4 seconds during which they were judged. The A and X elements were separated by pauses of 1 second.

### The PHONEME series

The second stimulus series (PHONEME series) was set up in the same way as the SYLLABLE series, except for two aspects. First, the second series was based on the disyllable triplet (4)-(6) instead of (1)-(3). Moreover, in the WORD+ condition of the second series the disyllables (4)-(6) were attached to the disyllabic noun *Cola* ([ˈkʰ<sup>hw</sup>o:l̩]), soft drink). It triggered a perceptual reinterpretation of the disyllables and led to the utterances (4) *Willst du mal + Cola* ‘do you want cola’, (5) *Haben wir + Cola* ‘Do we have cola’, (6) *Nehmen Sie + Cola?* ‘do you take cola?’. Unlike in the first series, all stimulus reinterpretations of the second series res-

ulted in the *same* number of syllables (i.e. 3 + disyllabic *Cola*). However, the numbers of segments/phonemes that can be restored with reference to the words emerging from the reinterpreted disyllables increased from (6) 2 over (5) 3 to (4) 5. The durations of the three stimuli in the ISO and WORD+ conditions of the second series were 620ms or 1.300ms, respectively. The 11 durations of the A elements in the two AX tests for the stimuli of the second series were adjusted accordingly. If the identification of the WORD+ stimuli (4)-(6) of the second series goes along with a phoneme-based restoration of the unreduced, canonical word forms, then the growing numbers of restored phonemes should increase the perceived length of the stimuli successively from (6) over (5) to (4). In fact, when produced in a non-reduced fashion the utterance (6) *Willst du mal Cola* is on average about 130ms longer than (5) *Haben wir Cola*, which, in turn, is about 70ms longer than the average utterance duration of (4) *Nehmen Sie Cola*. However, if there is restoration, but guided by syllables rather than phonemes, then the reinterpreted WORD+ stimuli (4)-(6) will sound similarly long. Finally, any kind of perceptual restoration will make the WORD+ stimuli appear longer relative to the A elements than the ISO stimuli of the control condition.

### Subjects and Performance

The 2x2 AX tests were cross-combined and presented in two separate perception experiments with different groups of 16 or 17 native speakers of German (12 male, 21 female subjects, 19-26 years old; gender was roughly balanced across the two groups). That is, group 1 received the two AX tests with the ISO stimuli of the SYLLABLE and with the WORD+ stimuli of the PHONEME series. Making the ISO vs. WORD+ variable a between-subjects factor obscured that the main focus of the experiment was on the subjects' (re-)interpretation and perception of the disyllables (1)-(6). The two AX tests of group 2 contained the ISO stimuli of the PHONEME and the WORD+ stimuli of the SYLLABLE series.

The two experiments took place subsequently in the same sound-treated room, in which the stimulus pairs were played over loudspeakers with constant sound settings. The stimulus pairs of each test were presented four times, twice as AX and twice as XA pairs, in an overall randomized order. Thus each test consisted of  $11 \times 3 \times 4 = 132$  stimulus pairs. An entire experiment had 264 pairs. The subjects were asked to compare the speech-like hum stimuli (A) with the actual speech stimuli (X) and to make a spontaneous judgment thereafter for each pair as to whether the speech-like stimulus or the speech stimulus was the longer one. Judgments were made by pressing buttons on a small box located in front of the subjects that also recorded the individual reaction times relative to the end of the stimulus pair.

After this introductory instruction, the subjects underwent a five-minute training session with pairs of the speech-like stimuli that showed the full spectrum of duration differences. The whole experiment had a duration of 1 hour, including a break between the 2 t test sessions of each experiment.

## Results

### General Remarks

Supplementary to the two perception experiments, an additional group of 12 native speakers of German was asked informally to reproduce the strings of words of the 2x6=12 ISO and WORD+ stimuli of the SYLLABLE and PHONEME series that were presented to them individually per loudspeaker. Some speakers reproduced the *Wissen + gucken* sequence as *Willst du ihn gucken* instead of *Willst du den gucken*. But this small variation does not affect the perceived syllable number. All other 11 ISO and WORD+ stimuli were reproduced with exactly the same wording as shown above. It can hence be assumed that these perceived wordings also underlie the subjects' relative length judgments in the actual experiments.

As regards the results of the actual experiments, it must at first be noted that the stimulus pairs showed a clear order-of-presentation effect. A t-test for paired samples was done that compared the frequencies of 'longer' judgments yielded by the A stimuli in all AX and XA pairs of the two experiments (i.e.  $n=264$  per sample). It was found that the A stimuli (constant speech-like hum with [ə] formants) were significantly more often deemed to be longer than the X stimuli (ISO disyllables or WORD+ utterances) in the XA order than in the AX order ( $t=-3.707$ ;  $df=263$ ;  $p<0.001$ ).

### Results of the SYLLABLE stimuli

Superimposed on the general effect of the second stimulus in the pair to sound longer than the first, which is also well-known from previous studies, the SYLLABLE series showed a striking effect of syllable number on the perceived duration of the X stimuli relative to the A stimuli. The results of the SYLLABLE stimulus pairs were analyzed by means of a series of 7 t-tests that were based on the frequencies of 'A longer X' judgments summed up across all subjects, which yielded values between 0 and 68. In terms of the concerned disyllables 6 t-tests dealt with the comparisons *Norma* vs. *Kermit*, *Norma* vs. *Wissen*, and *Kermit* vs. *Wissen* in both the ISO and the WORD+ conditions. Additionally, one t-test for independent samples (and heterogeneous variances, corrected by adjusting  $df$ ) compared the ISO condition (1) with the WORD+ condition (3), i.e. isolated *Norma* vs. *Wissen + gucken*. The findings are illustrated in Figure 1.

No significant differences in 'A longer X' judgments were found between the stimulus pairs of the ISO conditions (1)-(3). Therefore, all corresponding judgments were pooled in Figure 1. The resulting psychometric curve is characterized by quite an abrupt change from a clear (i.e. about 90%) 'X longer A' to an equally clear 'A longer X' perception that starts for the stimulus pairs with identical A and X durations, and that continues for the pairs in which the A stimuli were 75ms and 150ms longer than the X stimuli. The point of subjective equality (PSE) in terms of the perceived length of the A and X stimuli is +37ms. That is, in

the case of physically identical durations the X stimuli appeared slightly longer than the A stimuli. This perceptual bias may be due to the greater spectral and prosodic variation in the X stimuli, including a stimulus-final pitch fall which is known to induce perceptual lengthening (cf. Lehiste 1976).

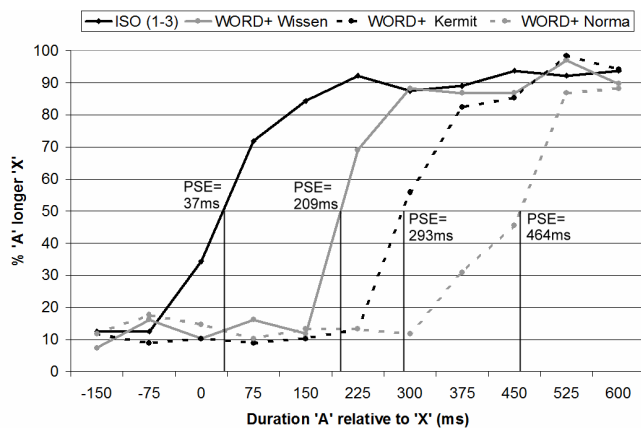


Figure 1: Percentages of ‘A longer X’ judgments yielded by the two AX tests based on the speech stimuli (X) of the SYLLABLE series when compared with the 11 speech-like hum stimuli (A) whose durations range from -150ms to +600ms relative to X. The data of the ISO condition were pooled across all disyllables (1)-3,  $n=192$  per percentage. The results of the WORD+ are presented separately for each utterance (1)-3,  $n=68$  per percentage.

Unlike in the ISO condition, the A stimuli in the WORD+ condition had to have durations of considerably more than 0-150ms in order to reach or exceed the perceived length of the X stimuli. The corresponding three judgment curves in Figure 1 illustrate how the perceptual changes to ‘A longer X’ occur likewise abruptly and as clearly as in the pooled ISO condition, but they required A durations of at least 225ms and up to 525ms. Moreover, the perceptual changes in the WORD+ conditions were successively delayed from *Wissen* over *Kermit* to *Norma*. Accordingly, the ‘A longer X’ frequencies decreased significantly in the corresponding t-tests from ISO to WORD+ *Wissen* ( $t=10.339$ ;  $df=31$ ;  $p<0.001$ ), and then within the WORD+ condition further from *Wissen* to *Kermit* ( $t=6.022$ ;  $df=16$ ;  $p<0.001$ ) and from *Kermit* to *Norma* ( $t=-6.473$ ;  $df=16$ ;  $p<0.001$ ), despite physically identical durations of all WORD+ stimuli. Of course, the comparison of the extreme stimuli *Wissen* and *Norma* was also highly significant ( $t=14.046$ ;  $df=16$ ;  $p<0.001$ ).

With regard to the points of subjective equality displayed in Figure 1 (i.e. 209ms, 293ms, 464ms), the results suggest as a rule of thumb that with every additional syllable that was identified in the wordings of the X stimuli the perceived length of X relative to A increased by 90-150ms.

Parallel to the syllable-related increase in the perceived length of the X stimuli an additional series of 7 t-tests revealed an increase in mean reaction times with a large step from ISO to WORD+ *Wissen* (459ms vs. 844ms;  $t=14.001$ ;

$df=31$ ;  $p<0.001$ ), and smaller steps from *Wissen* to *Kermit* (844ms vs. 937ms;  $t=8.506$ ;  $df=16$ ;  $p<0.001$ ) and from *Kermit* to *Norma* (937ms vs. 1266ms;  $t=-9.220$ ;  $df=16$ ;  $p<0.001$ ). In consequence, the t-test that compared *Wissen* and *Norma* also came out highly significant ( $t=17.227$ ;  $df=16$ ;  $p<0.001$ ). All other t-tests returned non-significant. A graphical summary of the mean reaction times of the SYLLABLE stimuli in the AX tests is provided in Figure 2.

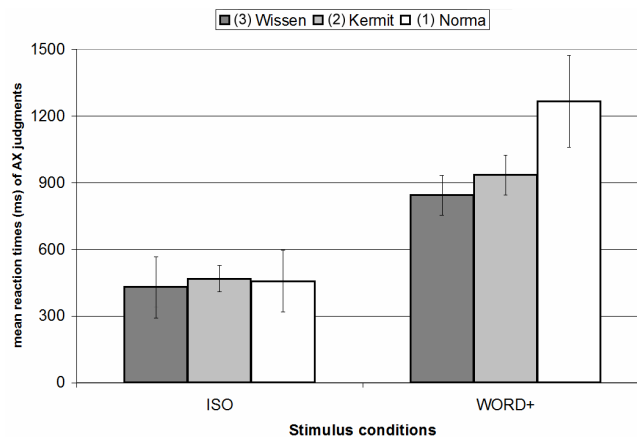


Figure 2: Reaction times (means and standard deviations in ms) of ‘A longer X’ judgments received for the speech stimuli (X) of the SYLLABLE series when compared with the 11 speech-like hum stimuli (A) in the two AX tests. Each bar represents 64 (ISO) or 68 (WORD+) judgments.

### Results of the PHONEME stimuli

The results of the AX tests based on the PHONEME stimuli are presented in Figures 3 and 4. Analogous to the results of the SYLLABLE series it can be seen in Figure 3 that each PHONEME stimulus X in both the ISO and WORD+ conditions led to a clear and rapid perceptual transition from ‘X longer A’ to ‘A longer X’ across the 11 comparisons with the increasingly longer A stimuli. Moreover, in terms of the majority of judgments all changes to ‘A longer X’ did not take place until the A stimuli showed much greater durations than the X stimuli. That is, the acoustically highly variable speech stimuli (X) were inherently longer than the acoustically constant hum stimuli (A). The PHONEME results also resemble the SYLLABLE results in that the durations needed for the A stimuli to be perceived as similarly long or longer than the X stimuli had to be significantly greater in the WORD+ than in the ISO condition (based on the comparison ISO (1) *Wismar* vs. WORD+ (3) *Nimm sie*:  $t=-13.812$ ;  $df=21$ ;  $p<0.001$  according to t-test for independent samples with heterogeneous variances and  $df$  correction). Consequently, the point of subjective equality of A and X lengths was located at about 66ms in the ISO condition and amounted to at least 188ms in the WORD+ condition. Finally, as in the SYLLABLE series, the greater perceived length of the PHONEME stimuli relative to A in the WORD+ condition went along with a substantial and also significant increase in mean reaction times (cf. Fig.4;  $t=-3.300$ ;  $df=26$ ;  $p=0.003$ ).

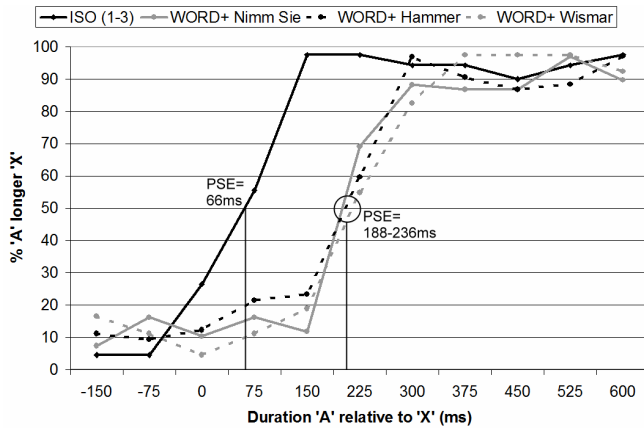


Figure 3: Percentages of ‘A longer X’ judgments yielded by the two AX tests based on the speech stimuli (X) of the PHONEME series when compared with the 11 speech-like hum stimuli (A) whose durations range from -150ms to +600ms relative to X. The data of the ISO condition were pooled across all disyllables (1)-(3),  $n=204$  per percentage. The results of the WORD+ are presented separately for each utterance (1)-(3),  $n=64$  per percentage.

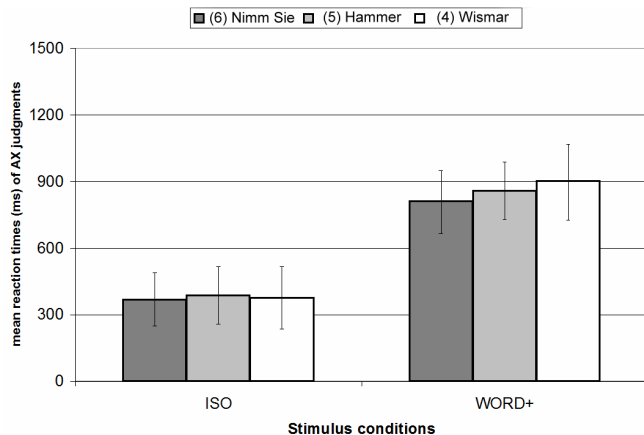


Figure 4: Reaction times (means and standard deviations in ms) of ‘A longer X’ judgments received for the speech stimuli (X) of the PHONEME series when compared with the 11 speech-like hum stimuli (A) in the two AX tests. Each bar represents 68 (ISO) or 64 (WORD+) judgments.

Apart from these similar results, the crucial difference between the SYLLABLE and the PHONEME stimuli is that the latter showed further significant effects within the WORD+ condition neither with regard to length judgments nor with regard to the corresponding reaction times. In terms of the points of subjective equality in relation to A, the lengths of the X stimuli differed by less than 50ms (188-236ms), as compared to more than 250ms (209-464ms, cf. Fig.2) in the SYLLABLE series. Similarly, the mean reaction times to the WORD+ PHONEME stimuli increased by less than 100ms from *Nimm sie* (812ms) to *Wismar* (902ms), whereas the parallel stimuli of the SYLLABLE series caused a reaction time increase of more than 400ms

(844-1266ms). In summary, the PHONEME disyllables *Wismar*, *Hammer*, and *Nimm sie* that were characterized by physically identical durations were also perceived as being equally long in the WORD+ stimuli, and they were processed homogeneously by the listeners in terms of reaction times, despite their different reinterpretations as *Willst du mal + Cola*, *Haben wir + Cola*, *Nehmen Sie + Cola*, and the different potentials for phonemic restorations that these reinterpretations offer.

## Conclusions

It was shown that the three sound patterns (1) [ $^1n\tilde{o}ʁ.m\ddot{a}$ ], (2) [ $^1k^h\ddot{a}ʃ.m^w\text{r}^t^h$ ], and (3) [ $^1v\ddot{u}s.s\ddot{a}n$ ] were identified by German listeners as the highly reduced function word sequences *Nun wollen wir mal*, *Können wir mit*, and *Willst du den/ihn* when followed by a semantic-prosodic trigger, i.e. the verb *gucken*. In unreduced forms these function word sequences have 1, 2, or 3 syllables more than the disyllabic nouns *Wissen*, *Kermit*, and *Norma*, which emerge from the sound patterns presented in isolation. This fact is perceptually reflected in successively increasing stimulus lengths (about 90-150ms per syllable) and reaction times (about 100-300ms per syllable) when judged in relation to a constant [ə]-like hum sound. These twofold judgments differences represent strong, though indirect, behavioral evidence that the function word sequences manifest themselves perceptually in fuller forms than just [ $^1n\tilde{o}ʁ.m\ddot{a}$ ], [ $^1k^h\ddot{a}ʃ.m^w\text{r}^t^h$ ], and [ $^1v\ddot{u}s.s\ddot{a}n$ ]. That is, the increase in syllable numbers was paralleled by restoration processes that took successively more time the more syllables were to be restored.

The same syllable-related restoration effect can be made responsible for the increases in perceived stimulus length and reaction time that occurred between the ISO condition (i.e. *Wismar*) and the WORD+ condition (i.e. *Nimm sie*) of the PHONEME series, since the function word sequences that emerged from (4) [ $^1v\ddot{u}s.m\ddot{a}$ ], (5) [ $^1h\ddot{a}m.m^j\ddot{a}$ ], and (6) [ $^1n\ddot{u}m^j.zi$ ] (*Nimm sie*) in this series show three syllables in their unreduced forms and hence one syllable more than the disyllables *Wismar*, *Hammer*, and *Nimm sie*. Of course, as in the SYLLABLE series the increase in reaction time from ISO to WORD+ in the PHONEME series could also partly be due to the Weber-Fechner law, i.e. the listeners’ sensitivities to duration differences and hence the confidence and velocity of their judgments decreased with the increasing physical stimulus duration that was given in the form of the attached *Cola* in the WORD+ condition. The attachment of *Cola* triggered the reinterpretation of *Wismar*, *Hammer*, and *Nimm sie* as *Willst du mal*, *Haben wir*, and *Nehmen Sie*. But although the different wordings that resulted from these reinterpretations offered the possibility to restore different numbers of phonemes (based on the corresponding canonical word forms), and although the different phoneme numbers are actually reflected in different average durations when the three reinterpreted utterances are produced in an unreduced fashion, there is no evidence from differences in relative perceived stimulus lengths or reaction times that

these phonemic or durational matters played a role in the listeners' judgments of the WORD+ stimuli. That is, the present findings do not give rise to the assumption that listeners made use of the possibility of restoring sub-syllabic, i.e. phonemic units in highly reduced speech.

This conclusion represents a major difference to Kemps et al. (2004:118) who found “*solid evidence for restoration of missing phonemes in reduced word forms*”. In this context it must be noted that the findings of the present study originated from an experimental task that avoided any explicit references to single sounds/phonemes or orthography/letters. Rather, the comparison of perceived stimulus durations was suitable to draw the subjects' attention away from the wordings of the stimuli. In contrast, Kemps et al. used orthographic representations and phoneme monitoring. From this point of view the present study supports previous assumptions according to which the empirical evidence in favour of the phoneme as the basic concept in spoken communication is at least partly due to the fact that the corresponding studies take the phoneme as their point of departure and hence inevitably create a phoneme bias. However, apart from this phoneme vs. syllable (or generally suprasegmental) issue, the present findings support that auditory restoration does occur in perceiving reduced speech, and that the cognitive processes that underlie this restoration manifest themselves indirectly in increased reaction times to the restored stimuli. The presence of a masker or interrupter, or the complete disappearance of units that are affected by restoration are no pre-requisites for this process.

Substantiating this initial empirical basis, follow-up studies must aim at shedding more light on the exact nature of the restoration, most importantly on the question *what* listeners actually restore. Do they restore only a kind of syllable-based durational grid with very vague phonetic content, or do they restore a phonetically much richer representation of the reduced sound patterns? Or can both be true depending, for example, on semantic and prosodic contexts or the cognitive load of the listener? For example, the fact that the increase in perceived stimulus length of about 90-140ms per syllable is well below the average syllable duration in German (Künzel 1987) could indicate that the restoration does not go all the way to the traditional canonical form. Furthermore, to what extent are the listeners' restorations, including the type of restored units (phonemes/syllables), task and stimulus specific, and what how do phonetic and phonological reference forms in the mental lexicons look like? The scope of these questions indicates that the main contribution of the present study was to (re-)open an important field of research rather than to provide definite and comprehensive answers.

### Acknowledgments

The author of this paper is particularly grateful to Laura Dilley, Meghan Clayards, Klaus Kohler, and the Marie-Curie Research Training Network “Sound 2 Sense” for all the lively discussions that gave the inspiration for this study. Moreover, thanks are due to Hartmut Pfitzinger and Declan Donaghey for their kind experimental and technical support.

### References

- Bashford, J.A. & Warren, R.M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, 42, 114-121.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341-345.
- Connine, C.M., Rambon, L.J., Patterson, D.J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70, 403-411.
- Hawkins, S. (1995). Arguments for a non-segmental view on speech perception. *Proc. XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden*, 18-25.
- Kemps, R., Ernestus, M., Schreuder, R. & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 117-127.
- Kohler, K.J. (1999). Articulatory prosodies in German reduced speech. *Proc. XIVth International Congress of Phonetic Sciences, San Francisco, USA*, 89-92.
- Kohler, K.J. & Niebuhr, O. (2011). On the role of articulatory prosodies in German message decoding. *Accepted for publication in Phonetica*.
- Künzel, H.J. (1987). *Sprechererkennung*. Heidelberg: Kriminalistik Verlag.
- Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4, 113-117.
- Mitterer, H. & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73-103.
- Niebuhr, O. & Kohler, K.J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Accepted for Publication in Journal of Phonetics*.
- Niebuhr, O. & Meunier, Ch. (2011). Sibilant-related vowel differences in French – Implications for the scope of assimilation. *Accepted for publication in Phonetica*.
- Ohalá, J.J. (1992). The segment – Primitive or derived? In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press.
- Samuel, A.G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1124-1131.
- Samuel, A.G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28-51.
- Shahin, A.J. & Miller, L.M. (2009). Multisensory integration enhances phonemic restoration. *J. Acoust. Soc. Am.*, 125, 1744-1750.
- Shinn-Cunningham, B. & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *J. Acoust. Soc. Am.*, 123, 295-301.
- Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393.
- Warren, R.M. & Sherman, G.L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, 16, 150-156.