Prosodic cues for morphological complexity:

The case of Dutch plural nouns

Rachèl J.J.K. Kemps, Mirjam Ernestus

Max Planck Institute for Psycholinguistics

Nijmegen, The Netherlands

Robert Schreuder

Interfaculty Research Unit for Language and Speech,

University of Nijmegen, The Netherlands

R. Harald Baayen

Interfaculty Research Unit for Language and Speech,

University of Nijmegen,

Max Planck Institute for Psycholinguistics,

Nijmegen, The Netherlands

Address all correspondence to: R.J.J.K. Kemps

Max Planck Institute for Psycholinguistics

P.O. Box 310

6500 AH Nijmegen

The Netherlands

Telephone: (+31) 24 3612160

Fax: (+31) 24 3521213

e-mail: rachel.kemps@mpi.nl

Abstract

It has recently been shown that listeners use systematic differences in vowel length and intonation to resolve ambiguities between onset-matched simple words (Davis, Marslen-Wilson, & Gaskell, 2002). The present study shows that listeners also use prosodic information in the speech signal to optimize morphological processing. The precise acoustic realization of the stem provides crucial information to the listener about the morphological context in which the stem appears, and attenuates the competition between stored inflectional variants. We argue that listeners are able to make use of prosodic information, even though the speech signal is highly variable within and between speakers, by virtue of the relative invariance of the duration of the onset. This provides listeners with a baseline against which the durational cues in vowel and coda can be evaluated. Furthermore, our experiments provide evidence for lexical storage of such prosodic information.

Prosodic cues for morphological complexity:

The case of Dutch plural nouns

Several studies in the visual modality have shown surface frequency effects in the comprehension of fully regular inflections, thus providing evidence for storage of the inflected form as a whole at some level of representation. These effects have been shown for both nouns and verbs, and in several languages. For regularly inflected verbs, evidence for full form storage has been found for Dutch (Baayen, Schreuder, De Jong, & Krott, 2002; Schreuder, De Jong, Krott, & Baayen, 1999) and for English (Alegre & Gordon, 1999). For regularly inflected nouns, evidence for full form storage has been found for Dutch (Baayen, Dijkstra, & Schreuder, 1997), Finnish (Bertram, Laine, Baayen, Schreuder, & Hyönä , 1999), English (Sereno & Jongman, 1997; Alegre & Gordon, 1999), and Italian (Baayen, Burani, & Schreuder, 1997).

Recently, experiments in the *auditory* modality have also shown effects of full form frequency for both nominal and verbal regular inflections in Dutch, suggesting the existence of full form representations of regularly inflected forms in the auditory modality as well (Baayen, McQueen, Dijkstra, & Schreuder, in press). This finding is surprising in the light of models of spoken word recognition that incorporate some form of lexical competition, such as the revised Cohort model (Marslen-Wilson, 1990; Marslen-Wilson, Moss, & Van Halen, 1996), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994). In these models, stored regularly inflected forms would be cohort competitors of their corresponding uninflected forms: In many languages (e.g., Dutch, German, English), the uninflected form is onset-embedded in the longer inflected form and thus, at the phonemic level, the signal is ambiguous until the offset of the last phoneme of the stem (e.g., uninflected (singular) form: *book*; inflected (plural) form: *books*). In other words, the two candidates will keep on competing for recognition (i.e., inhibiting one another) until after offset of the uninflected form. From this perspective, storage of regularly inflected forms seems dysfunctional. Storage of regularly inflected forms would create a recognition problem in the domain of inflection, similar to the recognition problem that exists outside the domain of inflection — as for example in the perception of onset-embedded words that have longer morphologically unrelated competitors, such as *ham* in *hamster*.

Using the frequency counts in the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn,

1993), we estimated how often both types of embedding (inflectional embedding versus morphologically unrelated embedding) occur in Dutch. We selected all 5129 monomorphemic lemmas that had a lemma frequency greater than zero. Subsequently, we selected all phonological word forms (uninflected and inflected forms) that corresponded to these lemmas. When we encountered a phonological form that contained an uninflected form at its onset, and that shared its stress pattern, we determined whether the stem of that form was the uninflected form. If so, we counted the phonological form as an inflectional continuation (e.g., [buk] - [bukə(n)], 'book' - 'books'). If the stem was not shared, we counted the phonological form as a morphologically unrelated continuation (e.g., [hɑm] - [hɑmstər], 'ham' - 'hamster'). This procedure resulted in the following counts: 2,188,144 tokens (307 types) were morphologically unrelated continuation forms, and 2,243,990 tokens (3015 types) were inflectional continuations. These numbers show that embedding is a highly frequently occurring phenomenon which challenges theories of word recognition. The documented existence of separate representations for uninflected and inflected forms makes models of lexical competition predict that listeners might be slowed down by lexical competition between onset-aligned candidates twice as often as when no storage of inflected forms would take place — unless listeners are somehow able to tell apart the phonemically ambiguous portions of the uninflected and the inflected forms.

In fact, evidence is accumulating that subtle subsegmental acoustic cues can reduce the ambiguity between onset-embedded words and their longer competitors, thus assisting the perceptual system in distinguishing them before the point in the acoustic signal at which disambiguating phonemic information comes in. Salverda, Dahan, and McQueen (in press) recorded participants' eye movements while they listened to Dutch sentences including a word with an onset-embedded word (e.g., *hamster* containing *ham*). The participants saw four pictures of objects on a computer screen and were instructed to use the computer mouse to move the picture of the object that was mentioned in the sentence. There were more fixations to a picture representing the embedded word (*ham*) when the first syllable of the target word (*hamster*) had been replaced by a recording of the embedded word than when it came from a different recording of the target word. Subtle acoustic information in the speech signal, namely, the duration of the embedded word relative to the duration of its corresponding syllable in the target word, appeared to lead the word-recognition system to favor the correct interpretation of lexically ambiguous spoken input.

Experiments by Davis et al. (2002) similarly suggest that both durational and intonational differences assist the perceptual system in distinguishing short words from longer morphologically unrelated words beginning with these shorter words. In a gating task, participants were presented with sentence fragments. In one condition (long-word condition), the sentence fragments ended in a long carrier word of which the initial syllable formed an onset-embedded word (e.g., *captain* containing *cap*). In the other condition (short-word condition), the sentence fragments ended in the short word corresponding to the intitial syllable of the carrier word followed by a word with an onset that matched the continuation of the longer carrier word (e.g., *cap tucked*, compare *captain*). The first syllable in the short-word condition was significantly longer than the first syllable in the long-word condition, and there was a marginally significant difference in average fundamental frequency (average fundamental frequency was higher in the long-word condition than in the short-word condition). Significantly more short-word responses were made to gates from short-word stimuli than to gates from long-word stimuli, suggesting that listeners are influenced by the acoustic differences that exist between short and long word sequences in responding to the initial syllables of the target word. Similar results were obtained in a cross-modal priming task. The stimuli from the gating task were presented up to the offset of the first syllable of the target word (e.g., *cap* from either *cap* or *captain*) as auditory primes, and were followed by a visual target that was either the short word (*cap*) or the long word (*captain*). Greater facilitation occurred when prime syllables came from the same word as the target.

Outside the domain of inflection, listeners thus appear to be sensitive to durational and intonational differences between short words and longer lexical competitors. It is not self-evident that such prosodic differences are also sufficiently present to be functional for inflected words. Consider the Dutch singular and plural forms of 'book': *boek* ([buk]) and *boeken* ([bukə(n)]).

First, two phonetic processes exert their influence in parallel: a *shortening* process and a *lengthening* process. For Dutch, the shortening process has been described by Nooteboom (1972). In a stress-timed language like Dutch, the duration of a stressed vowel reduces as a function of the number of unstressed syllables that follow (see also: Lehiste, 1972, and Fowler, 1977, for English; Lindblom & Rapp, 1973, for Swedish). Therefore, the duration of the vowel in the first syllable in *hamster* is expected to be shorter than the duration of the same vowel in *ham*. The same holds for the vowel in the first syllable in *boeken* as compared to the same vowel in *boek*. However, since the

second syllable in *boeken* is less complex than the second syllable in *hamster*, it is conceivable that the amount of shortening in words like *boeken* versus *boek* is smaller compared to the amount of shortening in words like *hamster* versus *ham*. The amount of shortening might not be enough to be functional for the listener.

Simultaneously, a prosodic lengthening process applies: The last syllable before a prosodic boundary (e.g., a prosodic word boundary or a phonological phrase boundary) is lengthened. Therefore, the form *ham* (which is followed by a word boundary) is expected to be longer than the first syllable in *hamster* (which is not followed by a word boundary). Cambier-Langeveld (2000) points out that when the rhyme of the last syllable consists of a schwa, as for example in words like *boeken* [bukə(n)], prosodic lengthening also applies to the penultimate syllable. In other words, in *hamster* only the last syllable is subject to prosodic lengthening, whereas in *boeken* both syllables are lengthened. Thus, it is likely that the difference between *boek* and the first syllable of *boeken* is smaller than the difference between *ham* and the first syllable of *hamster*.

Bearing these phonetic considerations in mind, it is not self-evident that durational modification of the first syllable occurs in inflected forms to the same extent as it does in words carrying onset-embedded morphologically unrelated words. The durational modification in inflected forms might not be sufficiently present to be functional.

Linguistic considerations lead to the same conclusion. Various linguists have argued that it is preferable for the phonological form of the stem to remain unaltered after affixation. For instance, Aronoff (1976) points out that affixes that leave their base words unchanged, i.e., that are phonologically transparent, are more productive than affixes that lead to phonological opacity (see also Dressler, Mayerthaler, Panagl, & Wurzel, 1987, for morphophonological processes in general). In Optimality Theory, this idea is implemented by means of alignment constraints between prosodic and morphological constituents (e.g., McCarthy & Prince, 1993). These linguistic considerations lead one to expect that it would be dysfunctional for the stem in isolation to differ from the stem followed by an inflectional ending.

Considered jointly, these phonetic and linguistic considerations show that it is not obvious that systematic subsegmental differences between inflected forms and their base words might exist and be functional for the listener.

On the other hand, the existence of functional prosodic differences in the domain of inflection

would resolve the paradox of the dysfunctionality of the storage of regular inflected forms in the auditory modality. Such differences would distinguish uninflected forms from their longer inflectional counterparts long before the offset of the uninflected form — their uniqueness point would then occur considerably earlier than suggested by their phonemic representation.

Interestingly, an indication that subsegmental differences may exist between uninflected forms and their longer inflectional counterparts is provided in a pilot study by Baayen et al. (in press). The singular and plural forms of five nouns were realised five times by four native speakers of Dutch. The mean duration of the singulars was longer (98 ms on average) than that of the stems embedded in the plurals.

The question arises whether such prosodic cues in the domain of inflection can be functional for the listener, given the enormous variability of speech within and across speakers. In the present paper, we address this question by means of an experimental study of regular plural nouns in Dutch. In Dutch, the regular plural form of many nouns consists of the noun stem and the plural suffix *-en* (usually realized as just a schwa; e.g., *boek* [buk] 'book' – *boeken* [bukə] 'books'). We studied both the combined and the independent effects of durational and intonational information in the speech signal on the processing of singular and plural forms, using a number decision task as well as a lexical decision task. We furthermore investigated whether such prosodic information might be part of stored lexical representations.

Experiment 1

The question addressed in Experiment 1, employing a number decision task, is whether listeners are sensitive to prosodic differences between singular forms and the stems of plural forms. If so, listeners are expected to be slowed down in their responses when there is a mismatch between the prosodic (durational and intonational) information in the acoustic signal of a word on the one hand, and the word's number as it is conveyed by the presence or absence of the plural suffix on the other hand. Moreover, the magnitude of the delay in response latencies is expected to covary with the degree of prosodic mismatch. We will test the covariance between degree of prosodic mismatch and magnitude of the delay in response latencies in a regression design. If listeners are not sensitive to prosodic differences between singular and plural forms, in other words, if listeners rely on segmental information only, mismatching prosodic information should not affect response

latencies.

Method

Participants. Forty-six participants, mostly students at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. From the CELEX lexical database (Baayen et al., 1993) we selected all Dutch monomorphemic nouns that met the following criteria: Their initial phoneme was not a vowel, their plural was formed by adding the suffix *-en* [ə(n)] to the stem, and they did not also function as verbal forms. Furthermore, they ended in an underlyingly voiceless plosive. In Dutch, the rule of final devoicing applies: Underlyingly voiced obstruents in syllable-final position are devoiced. The plural suffix *-en* [ə(n)] induces resyllabification of the stem-final obstruent as onset of the next syllable, and hence an underlyingly voiced stem-final obstruent will remain voiced (Booij, 1995). As a consequence, only stems ending in underlyingly voiceless obstruents phonemically have the same base in the singular as in the plural form. Therefore, we only selected nouns ending in an underlyingly voiceless plosive, so that no change of the voicing characteristics of the plosive would occur in the plural form. Finally, the singular surface frequencies and plural surface frequencies of the nouns were larger than zero. (Singular surface frequency and plural surface frequency are token counts. Token counts in CELEX are based on a corpus of 42 million words of written Dutch.) From the resulting group of 135 nouns, we selected 48 experimental nouns that contained a simplex coda. Additionally, we randomly selected 48 filler nouns from the group of 133 Dutch monomorphemic nouns that met all the above mentioned criteria, except that these nouns could also function as verbal forms.

We excluded nouns containing a complex coda for the following reason. As mentioned above, for stress-timed languages, the vowel duration in a stressed syllable decreases as a function of the number of unstressed syllables that follow (Nooteboom, 1972, for Dutch; Lehiste, 1972, and Fowler, 1977, for English; Lindblom & Rapp, 1973, for Swedish). This effect of the number of following syllables is smaller with smaller vowel duration in the stressed syllable (Nooteboom, 1972; Lehiste, 1972). In other words, the smaller the vowel duration in the monosyllabic singular form, the smaller the difference that is to be expected between the vowel duration in the singular

form and the vowel duration in the bisyllabic plural form. Since vowels have a smaller duration when they are followed by a complex coda than when they are followed by a single consonant (Waals, 1999, for Dutch; Munhall, Fowler, Hawkins, & Saltzman, 1992, for English), the difference between singular and plural forms is expected to be smaller for words ending in a complex coda than for words ending in a single consonant. We decided to exclude nouns with a complex coda, so that the durational difference to be expected between the vowel in the singular form and the vowel in the plural form was maximal.

Three reading lists were created: a list containing the singular forms of the experimental nouns, a list containing the plural forms of the experimental nouns, and a list containing the plural forms of the filler nouns. The order of the nouns within lists was randomized three times, resulting in 9 reading lists. In order to maximize durational differences between singular and plural forms, the noun forms were read in isolation. The lists were recorded in a soundproof recording booth by a native female speaker of Dutch, who was naive regarding the purpose of the experiment. The recordings were digitised at 16 kHz.

For each noun form, the best realisation (of three) was selected and spliced out of its list using the PRAAT speech editing software (Boersma, 1996). Subsequently, out of the experimental noun forms we created two types of singular forms: 'normal' singular forms and 'constructed' singular forms. The normal singular form consisted of the singular form exactly as it was uttered by the speaker. The constructed singular form consisted of the stem of the plural form — in other words, it was the plural form with the plural suffix *-en* [ə(n)] spliced off. The point of splicing was located at the onset of the voicing of the schwa following the stem-final consonant. As a result, the normal singular form's prosodic information matched its number information (as conveyed by the absence of the plural suffix), whereas the constructed singular form's prosodic information mismatched its number information: It's prosodic characteristics signalled a plural form, whereas the absence of the plural suffix signalled a singular form. Total duration, vowel duration, closure duration, and release noise duration were measured for the two types of singular forms. Onset of the vowel was defined as onset of voicing if the preceding segment was voiceless, and as the end of the release noise if the vowel followed a fully voiced stop. In all other cases (i.e., if the preceding segment was [l, ʀ, m, n, v] or [ʋ]), onset of the vowel was defined as the point of change in the periodicity pattern in the waveform. The end of the vowel and beginning of closure was

defined as the end of the second formant of the vowel. The end of closure was located at the onset of the sudden discontinuity in the waveform for the release noise. A paired t-test showed that on average the constructed singular forms had a significantly smaller total duration than the normal singular forms ($t(47) = 18.2, p < 0.0001$). The mean difference in total duration was 87 ms. The mean difference in vowel duration was 18 ms ($t(47) = 14.8, p < 0.0001$), the mean difference in closure duration was 26 ms ($t(47) = 10.9, p < 0.0001$), and the mean difference in release noise duration was 37 ms ($t(47) = 13.8, p < 0.0001$). An analysis of variance with total duration as the dependent variable showed that there was no significant interaction between type of singular form (normal versus constructed singular form) and type of vowel (phonologically and phonetically long versus short vowel): The difference in duration between normal and constructed singular forms was comparable for words with phonologically and phonetically long and short vowels ($F(1, 92) = 0.4, p = 0.52$). Furthermore, we measured the average fundamental frequencies of the normal and the constructed singular forms. Recall that Davis et al. (2002) found that the average fundamental frequency was higher in the initial syllables of bisyllabic words than in monosyllabic words. We obtained a similar result: The constructed singular forms had a significantly higher average fundamental frequency than the normal singular forms ($t(47) = -2.0, p < 0.05$). The mean difference in average fundamental frequency was 7 Hz. Our explanation for this finding is that all (monosyllabic and bisyllabic) forms were pronounced with an intonational phrase final pitch accent H*L, which was aligned differently in monosyllabic than in bisyllabic words. In the case of the monosyllabic forms, both H and L were realized within one syllable. In the case of the bisyllabic forms, H was assigned to the first (stressed) syllable and L was assigned to the second syllable. Consequently, average fundamental frequency was higher in the first syllables of the bisyllabic forms than in the monosyllables. Table 1 lists the mean durations with their standard deviations and the mean average fundamental frequencies with their standard deviations for the two kinds of singulars.

<center>PLACE TABLE 1 APPROXIMATELY HERE</center>

The normal and constructed singular forms functioned as experimental target words. Filler words consisted of the plural filler nouns, exactly as they were uttered by the speaker.

Three experimental trial lists and their complements were created in such a way that each list

contained all 48 filler items, 24 normal singular forms, and 24 constructed singular forms. One list never contained both the normal and the constructed singular form of a single noun: If a given list contained the normal singular form of a noun, then the constructed singular form of that noun was contained in its complementary list. The order of presentation of the stimuli was pseudo-randomized within the three lists: No more than three singular forms of the same type occurred successively. Orders were identical in complementary lists. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 8 plural forms, 4 normal singular forms, and 4 constructed singular forms (all taken from a different recording of the complete experimental and filler sets). None of the nouns in the practice set was presented in the actual experiment.

Procedure. Participants were instructed to decide as quickly as possible whether the form they heard was a singular or a plural form. They responded by pressing one of two buttons on a button box. All experimental items required the response 'singular', whereas all filler items required the response 'plural' (assuming that decision on number is primarily based on the presence c.q. absence of a plural suffix). Each trial consisted of the presentation of a warning tone (377 Hz) during 500 ms, followed after an interval of 450 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 10 minutes.

Results and discussion

No participants or items were excluded from the analyses, since they all showed error rates below 20%. In all analyses, only trials eliciting correct responses were included. The mean reaction times (measured from word offset) for the two kinds of singulars are summarized in Table 2.

PLACE TABLE 2 APPROXIMATELY HERE

If listeners are sensitive to prosodic differences between singular and plural forms, our depen-

dent variable reaction time should covary with the degree of prosodic mismatch between normal and constructed singular forms. Simply finding a delay in processing (109 ms; $t1(45) = -16.0, p < 0.0001; t2(47) = -15.4, p < 0.0001$) is not sufficient evidence for the occurrence of a prosodic mismatch effect, as this delay might as well be a consequence of the splicing manipulation that has been applied to the constructed singular forms. What needs to be shown is a correlation between the magnitude of the prosodic mismatch and the delay in processing.

We therefore applied a covariance analysis along the lines of Lorch and Myers (1990) to the reaction time data corresponding to the constructed singular forms. We operationalised the amount of prosodic mismatch as the difference between the duration of the constructed singular form and the duration of the corresponding normal singular form. As mismatch in intonational contour is not straightforwardly quantifiable — average fundamental frequency does not capture contour information — we did not include intonational mismatch in the numeric operationalisation of prosodic mismatch. It is conceivable, however, that the amount of intonational mismatch codetermined reaction times to the constructed singular forms, and we will return to this issue below. We fitted a linear model to the data for each participant separately, with log reaction time as the dependent variable, and log singular surface frequency, duration of the form itself, and the durational difference score as predictors. Singular surface frequency was included as a predictor in order to ascertain that any observed effect of the durational difference score could not be a consequence of confounding durational differences with differences in frequencies between the items. T-tests on the coefficients of the participants for the predictors revealed that duration had a facilitatory effect (the longer the duration, the shorter the response latencies; $t(45) = -3.0, p < 0.01$), and durational difference had an inhibiting effect (the larger the durational mismatch, the longer the response latencies; $t(45) = -3.0, p < 0.01$).

Using a multi-level extension of the Lorch and Myers-technique (Pinheiro & Bates, 2000), we tested whether any effect of durational difference remained after partialling out the effects of the other predictors. This was indeed the case ($F(1, 1035) = 6.0, p < 0.05$), indicating that durational difference had an independent effect on the reaction times to the constructed singular forms.

Apparently, when listeners segmentally perceive a singular form, but prosodically (durationally) a plural form is signalled, their number decision is adversely affected. What then happens in the opposite situation? What happens when segmentally a plural form is presented, but prosodic

cues in the stem signal a singular form? In Experiment 2 we investigated whether we may replicate this prosodic mismatch effect for plural forms. We created two types of plural forms: one form in which the prosodic (durational and intonational) cues *matched* the number of the form as it was conveyed by presence of the suffix, and one form in which the prosodic cues *mismatched* the number of the form as conveyed by the presence of the suffix.

Experiment 2

Method

Participants. Forty-three participants, mostly students at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 1.

Materials. The target items in this experiment were normal and constructed *plural* forms. Contrary to in Experiment 1, both types were now created by means of a splicing manipulation, which allows a factorial experimental design contrasting normal and constructed forms. The filler items were now singular forms.

We selected the same experimental and filler nouns as in Experiment 1. The singular forms of the experimental nouns, the plural forms of the experimental nouns, and the singular forms of the filler nouns were assigned to separate reading lists. The order of the nouns within lists was randomized three times, resulting in 9 reading lists. These lists were read by the same native female speaker as in Experiment 1. The lists were recorded in a soundproof recording booth. The recordings were digitised at 16 kHz.

Subsequently, we created the two types of plural forms: normal plural forms and constructed plural forms. Both types of plural forms were created using a splicing technique: The beginning of one speech token was combined with the ending of a different speech token. From both the singular and the plural form of a noun, we selected the portion of signal from the first phoneme up to and including the closure of the final plosive of the stem. In other words, we selected the stem without the release noise of the final plosive. From another realisation of the plural form of the same noun, we selected the portion from the release noise of the final plosive of the stem up to and

including the last phoneme. To create the normal plural form, we concatenated this latter portion to the initial portion originating from the plural from. To create the constructed plural form, we concatenated it to the initial portion originating from the singular form. As a result, the normal plural form consisted of two portions of signal both originating from plural forms, whereas the constructed plural form consisted of an initial portion originating from a singular form and a final portion originating from a plural form. Note that by applying this splicing procedure to both the normal and the constructed plural forms, we ensured that any observed difference in response latencies can not be a consequence of a difference in splicing manipulation: A delay in processing for the constructed plural forms would constitute sufficient evidence for the occurrence of a prosodic mismatch effect.

Since the initial portion of the constructed plural form originated from a singular form, it was expected to contain durational and intonational cues that mismatched the number of the word as it was conveyed by the presence of the plural suffix. A paired *t*-test indeed showed a significant difference in total duration between the normal and the constructed plural form: The constructed plural form was longer (29 ms on average) than the normal plural form ($t(47) = 5.6, p < 0.0001$). The difference in vowel duration was 16 ms ($t(47) = 6.6, p < 0.0001$) and the difference in closure duration was 19 ms ($t(47) = 6.4, p < 0.0001$). In addition, intonational differences were present between the initial portions of the normal and constructed forms: The average fundamental frequency of the initial portion of the constructed plural form was on average 11 Hz lower than the average fundamental frequency of the initial portion of the normal plural form ($t(47) = -10.9, p < 0.0001$). Table 3 lists the mean total durations with their standard deviations and the mean average fundamental frequencies of the initial portions with their standard deviations for the two types of plural forms.

PLACE TABLE 3 APPROXIMATELY HERE

In the case of the constructed plural forms, this splicing procedure gave rise to artificial plosives that combined the closure of a singular realisation with the release noise of a plural realisation. Or, put differently, durational information contained in the original release noise of the singular realisation was no longer present in the acoustic signal of the constructed plural form. Recall that we applied this splicing procedure in order to ensure that any observed difference in

response latencies can not be a consequence of a difference in splicing manipulation. But would it have been more natural, and more analogous to the creation of the constructed singular forms in the previous experiment, to simply concatenate the plural suffix to the singular stem when forming constructed plural forms? Actually, it turned out that this latter procedure gave rise to very unnatural sounding stimuli. In fact, this by itself already exactly answers our research question: A plural form is not simply a singular form with a plural suffix concatenated to it, neither in production nor in perception. The stem in the plural form contains acoustic cues that distinguish it from the same stem in the singular form. In order to prevent participants from showing unnatural behavior as a result of the presence of very unnatural sounding stimuli in the experiment, and in order to determine whether prosodic cues other than the nature of the release noise play a role in the processing of plurals, we opted for the present, somewhat more complicated cross-splicing procedure.

Three trial lists and their complements were created in the same manner as in Experiment 1: Each list contained all 48 filler items, 24 normal plural forms, and 24 constructed plural forms. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the experiment. The practice set consisted of 16 trials: 8 singular forms, 4 normal plural forms, and 4 constructed plural forms. None of the nouns in the practice set was presented in the actual experiment.

<u>Procedure.</u> The same experimental procedure was used as in Experiment 1, except that now all experimental items required the response 'plural' and all filler items required the response 'singular' (again assuming that number decision is primarily based on the presence c.q. absence of a plural suffix).

Results and discussion

All items and subjects were included in the analyses, since they all showed error rates below 20%. Table 4 lists the mean reaction times to the two types of plural forms.

PLACE TABLE 4 APPROXIMATELY HERE

Paired *t*-tests showed a significant difference in response latencies: Response latencies to the

constructed plural forms were longer (24 ms on average) than to the normal plural forms ($t1(42) = -3.6, p < 0.001; t2(47) = -2.3, p < 0.05$). The physically longer items thus produced the longer reaction times. A simple processing explanation (i.e., longer signal to process, longer processing time), however, seems rather unlikely, since reaction times were measured from word offset. Furthermore, the covariance analysis in Experiment 1 shows that duration in fact has a facilitatory effect: longer item durations were associated with shorter reaction times.

Instead, the prosodic mismatch effect originally observed for singular forms appears to have occurred for plural forms as well. Interestingly, the effect for the plural forms was considerably smaller than the effect for the singular forms (24 ms for the plurals in Experiment 2 as opposed to 109 ms for the singulars in Experiment 1). There are two possible explanations for this.

First, the magnitude of the prosodic mismatch was larger for the singulars in Experiment 1 than for the plurals in Experiment 2. Whereas in Experiment 1 all durational information carried by the stem of the plural form was contained in the constructed singular form, in Experiment 2 durational information contained in the release noise of the final plosive of the singular was no longer present in the constructed plural form as a consequence of the splicing procedure. An explanation of the different delay magnitudes between experiments in terms of different mismatch magnitudes is supported by a comparison of the correlations between durational differences and reaction time differences in the two experiments. In Experiment 1, a significant correlation was observed (Spearman's $\rho = -0.34, p < 0.05$). In Experiment 2, in which the durational differences were considerably smaller, this correlation was marginally significant (Spearman's $\rho = -0.19, p < 0.10$, one-tailed) but was not significantly different from the correlation in Experiment 1 ($Z = 0.8, p = 0.44$). In other words, 'the larger the durational difference, the larger the difference in reaction times' seems to apply both within and across experiments.

Second, the nature of the expectancy violation in Experiment 1 was different from that in Experiment 2. In Experiment 1, presentation of the constructed singular form led the listener to expect a plural form on the basis of the durational (and possibly intonational) cues that were present in the signal, but then suddenly the auditory signal was broken off, leaving the listener with conflicting evidence. In Experiment 2, presentation of the constructed plural form initially led the subjects to expect a singular form, but then the auditory signal continued until the end of the plural suffix. Evidence pointing to the plural form thus kept accumulating after the stem,

partly compensating for the subtle prosodic cues in the stem pointing to the singular form. It is possible that this difference in the nature of the violation of the expectancy was also reflected in the different magnitudes of the prosodic mismatch effect in response latencies.

The covariance analysis described under Experiment 1 showed that reaction times to the constructed singular forms in that experiment were at least partly determined by the magnitude of the durational mismatch between the normal and the constructed forms. As mentioned before, mismatch in intonational contour is not as easily quantifiable, and can therefore not similarly be included as a predictor in a linear model. We therefore investigated the individual contribution of intonational information to the prosodic mismatch effect in a separate experiment. In Experiment 3, again, normal and constructed singular forms were presented, but now these two types of singular forms only differed in intonational contour. If intonational cues contribute to the prosodic mismatch effect, we should observe longer response latencies to the forms with the mismatching intonational contour.

Experiment 3

Method

Participants. Forty-nine participants, mostly students at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 1 or 2.

Materials. The normal singular forms from Experiment 1 were used with no further manipulation. In addition, new constructed singular forms were created by taking the normal singular forms, and overlaying them with the intonational contours taken from the stems of the plural forms. This manipulation was carried out using the PSOLA (Pitch-Synchronous Overlap and Add) resynthesis method in the PRAAT speech editing program (Boersma, 1996). As a result, the durations of the two types of singular forms were identical, but one type of singular form carried the intonational contour of the singular ('normal' singular form), whereas the other type of singular form carried the intonational contour of the plural ('constructed' singular form). The same filler words were used as in Experiment 1.

Three trial lists and their complements were created in the same manner as in the previous experiments: Each list contained all 48 filler items, 24 normal singular forms, and 24 constructed singular forms. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 8 plural forms, 4 normal singular forms, and 4 constructed singular forms. None of the nouns in the practice set was presented in the actual experiment.

Procedure. The same experimental procedure was followed as in the previous experiments.

Results and discussion

We included all items and participants in the analyses, since they all showed error rates below 20%. Table 5 lists the mean reaction times to the two kinds of singular forms.

PLACE TABLE 5 APPROXIMATELY HERE

Participants responded on average 10 ms slower to the constructed singular forms than to the normal singular forms. In a paired *t*-test, this difference was significant by subjects ($t1(48) = -2.2, p < 0.05$), but not by items ($t2(47) = -1.5, p = 0.14$). As the item-analysis may be too conservative for the type of experimental design used in this study (Raaijmakers, Schrijnemakers, & Gremmen, 1999), we additionally ran a covariance analysis (Lorch & Myers, 1990), in which the factor condition (normal singular form versus constructed singular form) and the covariate log singular surface frequency predicted log reaction times. This analysis revealed significant effects of both condition ($t(48) = -2.8, p < 0.01$) and of singular surface frequency ($t(48) = -4.7, p < 0.0001$).

These results show that when intonational information mismatches number information (conveyed by the presence/absence of the plural suffix), number decision is hindered. Both duration and intonation thus appear to serve as cues in perceptually distinguishing between singular and plural forms. The processing delay for stimuli with mismatching intonational contour was only 10 ms. Note, however, that the stimuli in our experiments were presented in isolation. The participants did not hear surrounding speech that could function as a frame of reference against which they could evaluate the fundamental frequency of the stimuli. It is conceivable that, when singu-

lars and plurals are presented in their context, intonation serves as a considerably stronger cue than it did in this experiment. An alternative explanation for the relatively small effect of intonational mismatch on reaction times is that the intonational difference is peculiar to the context in which the words were produced — contrary to the durational difference, which is probably quite systematically present between singulars and plurals produced in any context. In a list context, each word will have an intonational phrase final contour. This contour will be aligned differently for monosyllables than for bisyllabic forms, leading to differences in average fundamental frequency in the first syllable. However, singulars and plurals do not typically occur in phrase final position, and will therefore not show differences in average fundamental frequency as systematic as the durational differences. If intonational differences are indeed less systematic than durational differences, it is not surprising that listeners are less sensitive to intonational mismatch than to durational mismatch.

In Experiment 4, we addressed the question of whether there are other acoustic cues in addition to and independent of duration and intonation, that might help the perceptual system in distinguishing between the singular form and the stem of the plural form. For instance, the stem-final plosive in the constructed singular form may contain coarticulatory information about the following vowel, which is absent in the normal singular form. However, as the following vowel is a schwa, which has a very neutral articulation, we expect — if any — only weak coarticulatory cues. We presented normal and constructed singular forms that did not differ in duration nor in intonation, but that may differ in coarticulatory cues. If no cues other that duration or intonation have a perceptual effect, response latencies should not differ in Experiment 4.

Experiment 4

Method

Participants. Forty participants, mostly students at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiments 1 to 3.

Materials. In order to obtain normal singular forms and constructed singular forms (i.e., stems

of plural forms) that showed minimal intonational and durational differences, we embedded the singular and plural forms of all 135 nouns from the complete experimental set and the plural forms of all 133 nouns from the complete filler set (see Experiment 1) in a frame sentence in which the noun carried sentence accent: *Ik raak de/het X aan.* ('I touch the X.'). The sentences were assigned to separate reading lists for experimental nouns and for filler nouns. In the reading list containing the experimental nouns, the sentence containing the singular form of a particular noun was always followed by the sentence containing the plural form of that noun. By means of constructing the reading list in this way, we hoped to bring the speaker in a speaking rhythm such that intonational and durational differences between singular forms and the stems of plural forms were minimized. The reading lists were recorded twice in a soundproof recording booth. The speaker was the same as in Experiment 1 and 2. The recordings were digitised at 16 kHz.

Subsequently, the best realisations (of two) of the singular forms of the experimental nouns (normal singular forms), of the stems of the plural forms of the experimental nouns (constructed singular forms), and of the filler nouns (plural forms) were spliced out of their context. We selected those experimental nouns which did not show substantial differences in intonational contour between the normal and the constructed singular form (established both visually and auditorily). This procedure resulted in a set of 40 nouns. A paired t-test showed that within this set, no significant differences in duration existed between the normal and the constructed forms $(t(39) = -1.0, p = 0.34)$. As filler words, 40 (plural) nouns were randomly selected from the filler set.

Three trial lists and their complements were created in the same manner as in the previous experiments: Each list contained all 40 filler items, 20 normal singular forms, and 20 constructed singular forms. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the experiment. The practice set consisted of 16 trials: 8 plural forms, 4 normal singular forms, and 4 constructed singular forms. None of the nouns in the practice set was presented in the actual experiment.

<u>Procedure.</u> The same experimental procedure was followed as in the previous experiments.

<u>Results and discussion</u>

All items were included in the analyses, since they all elicited error rates below 20%. The data of three participants were excluded from the analyses, since they performed with overall error rates above 20%. Table 6 lists the mean reaction times to the two kinds of singular forms.

PLACE TABLE 6 APPROXIMATELY HERE

The mean difference in reaction times was 11 ms, with normal singulars eliciting longer response latencies instead of shorter response latencies. Paired *t*-tests showed that this difference was not significant ($t1(36) = 1.2, p = 0.25$; $t2(39) = 1.0, p = 0.32$).

Note that in this experiment, even though the two types of singular forms had undergone a different splicing manipulation (splicing within the word for the constructed singular forms versus no splicing within the word for the normal singular forms), there was no effect in the reaction times. In Experiment 1, we thought it necessary to show that there was a correlation between the magnitude of the delay and the amount of prosodic mismatch, as it seemed conceivable that not only the prosodic mismatch, but also the splicing manipulation had adversely affected the reaction times to the constructed forms. The null result in the present experiment suggests that splicing may not have been as problematic as we thought, and that the observed delay in processing times in Experiment 1 by itself already constitutes valid evidence for the prosodic mismatch effect.

To conclude, the results of the present experiment indicate that, apart from durational and intonational cues, the acoustic signal does not contain other, independent prosodic cues that might help the perceptual system in distinguishing between singular forms and the stems of plural forms: The relevant prosodic information in the signal consists of durational and intonational cues (or cues that covary with durational or intonational cues).

In all experiments so far we employed a number decision task. In the next and last experiment, we replicate the basic finding using another experimental paradigm, auditory lexical decision. We opted for lexical decision for two reasons. First, auditory lexical decision is a task in which the number of syllables is irrelevant: Whereas for number decision the number of syllables, and thus the prosodic structure of the stem, is informative with respect to the decision to be made, for lexical decision it is not. A first question addressed by Experiment 6 therefore is whether listeners are also sensitive to prosodic cues under these circumstances. Second, the responses to normal and constructed pseudoword singulars will shed light on whether the prosodic mismatch effect

observed for existing words results purely from the representations stored in the mental lexicon or whether it is mediated at some prelexical level.

Experiment 5

Method

Participants. Forty-two participants, mostly students at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiments 1 to 4.

Materials. Four experimental item types were included in the experiment: normal and constructed singular *word* items, and normal and constructed singular *pseudoword* items. The word items were the exact experimental items as used in Experiment 1 (i.e., 48 normal singular forms and 48 constructed singular forms).

Out of the singular word items, 48 singular pseudoword items were created by changing one to three phonemes in such a way that the phonotactic constraints of Dutch were not violated, and that the pseusowords' prosodic structure was identical to that of the words. Subsequently, the 'plural' forms of these pseudowords were created by adding the plural suffix *-en* [ə(n)], which is the appropriate allomorph as the stems consist of a single syllable. The 48 singular and 48 plural forms were assigned to separate reading lists. The orders within these lists were randomized twice, resulting in 4 reading lists.

Additionally, 100 filler words were included in the experiment: 25 monomorphemic, uninflected nouns, 25 inflected nouns (plural and diminutive inflections), 25 uninflected and inflected verbs, and 25 uninflected and inflected adjectives. The number of syllables of the filler words ranged from one to three. Out of these filler words, 100 filler pseudowords were created by changing one to three phonemes, again in such a way that the phonotactic constraints of Dutch were not violated, and that the pseudowords' prosodic structure was identical to that of the words. The filler words and the filler pseudowords were assigned to one reading list. The order within this list was randomized three times, resulting in three reading lists.

One more reading list was created consisting of 10 words, 5 'singular' pseudowords, and 5

'plural' pseudowords. These items were used to create practice trials. The order within this list was randomized twice, resulting in two reading lists.

All 9 reading lists were recorded by the same native female speaker of Dutch as in the previous experiments. The recordings were made in a soundproof recording booth and subsequently digitised at 16 kHz.

From the reading lists containing the experimental pseudoword items, the best realisations (of two) of the singular and the plural forms were selected. The singular forms served as the normal singular pseudoword items. Constructed singular pseudoword items were created by splicing the 'stems' out of the plural forms. From the reading lists containing the filler items, the best realisations (of three) of all filler words and of all filler pseudowords were selected. Finally, from the lists containing the practice items, the best realisations (of two) of all words and pseudowords were selected. The 10 words and the 5 singular pseudowords were spliced out of the lists exactly as they were realised by the speaker. From the plural pseudowords we only selected the stems, creating 5 constructed singular pseudowords.

Three experimental trial lists and their complements were created in such a way that a given list contained 100 filler words, 100 filler pseudowords, 24 normal singular word items, 24 constructed singular word items, 24 (or 23) normal singular pseudoword items, and 23 (or 24) constructed singular pseudoword items. (Due to an error, one pseudoword item had to be removed from the design). One list never contained both the normal and the constructed singular form of a single noun (word or pseudoword): If one list contained the normal singular form of a noun, then the constructed singular form of that noun was contained in its complementary list. The order of presentation of items was pseudo-randomized: No more than three singular forms of the same type occurred successively. Orders were identical in complementary lists. Participants were randomly assigned to experimental trial lists. The twenty practice trials were presented prior to the experiment.

The pseudoword items showed differences in duration and in intonation, similar to those observed in the word items (see Experiment 1): The normal singular forms were significantly longer (95 ms on average) than the constructed singular forms ($t(46) = 21.5, p < 0.0001$), and the constructed singular forms had a higher average fundamental frequency (5 Hz on average) than the normal singular forms ($t(46) = -1.9, p < 0.1$). The mean difference in vowel dura-

tion was 15 ms ($t(46) = 8.1, p < 0.0001$), the mean difference in closure duration was 16 ms ($t(46) = 6.0, p < 0.0001$), and the mean difference in release noise duration was 70 ms ($t(46) = 26.1, p < 0.0001$). Table 7 lists the mean durations and the mean average fundamental frequencies for normal and constructed singular words and pseudowords. The differences between the normal and the constructed forms were not significantly different for words than for pseudowords (duration: $F(1, 93) = 1.5, p = 0.22$; fundamental frequency: $F(1, 93) = 0.2, p = 0.64$).

PLACE TABLE 7 APPROXIMATELY HERE

Procedure. Participants were instructed to decide as quickly as possible whether the form they heard was a word or a pseudoword. They responded by pressing one of two buttons on a button box. Each trial consisted of the presentation of a warning tone (377 Hz) during 500 ms, followed after an interval of 450 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. Two short pauses were included in the experiment, resulting in three experimental trial blocks of approximately equal size. The total duration of the experimental session was approximately 30 minutes.

Results and discussion

The data of all participants were included in the analyses, since they all showed error rates below 20%. Nine word items and three pseudoword items elicited error rates above 20%. These items and their corresponding forms in the complementary condition were excluded from the analyses. Table 8 lists the mean reaction times to the four experimental item types.

PLACE TABLE 8 APPROXIMATELY HERE

The reaction times to the constructed experimental forms were significantly longer (74 ms on average) than the reaction times to the normal experimental forms ($F1(1, 41) = 100.4, p <$

$0.0001; F2(1, 81) = 55.6, p < 0.0001$). Given the null result in Experiment 4, this already suggests listeners' sensitivity to prosodic mismatch. In order to completely rule out the possibility that the observed delay to the constructed singular forms is the result of the splicing manipulation applied to these forms, we ran a covariance model on the reaction time data for the constructed singular forms. A linear model was fitted to the data of each participant separately (cf. Lorch & Myers, 1990), in which log reaction times were predicted by the duration of the form itself, by the durational difference score (i.e., the difference in duration between the normal and the constructed form), and by lexical status (word versus pseudoword). T-tests on the coefficients of the subjects on the three predictor variables yielded main effects of all three predictors (duration: $t(41) = -8.6, p < 0.0001$; durational difference: $t(41) = 2.8, p < 0.01$; lexical status: $F(1, 41) = -8.2, p < 0.0001$). Duration had a facilitatory effect: the longer the duration, the shorter the response latencies. Interestingly, there was a significant interaction between durational difference and lexical status ($t(41) = -4.9, p < 0.001$). A multi-level extension of the Lorch and Myers technique (Pinheiro & Bates, 2000) revealed that for words, durational difference had an inhibiting effect: the larger the durational difference, the longer the reaction times ($t(1815) = -3.2, p < 0.01$). For pseudowords, however, we obtained the opposite effect: the larger the durational difference, the shorter the reaction times ($t(1815) = 2.11, p < 0.05$). In other words, large prosodic (durational) mismatch appears to make words less word-like and pseudowords more pseudoword-like. A comparison between words and pseudowords of the coefficients for the correlation between durational differences and reaction times revealed that this correlation was significantly stronger for words than for pseudowords ($Z = -2.3, p < 0.05$). This stronger effect of prosodic mismatch for words than for pseudowords suggests that prosodic information is part of the lexical representations for words. For both words and pseudowords, the expectation regarding the syllabic structure of the word is based on generalizations over all singulars and plurals stored in the mental lexicon. For words, however, there is additional lexical support for this expectation, as the word itself, its syllabic structure and the prosodic characteristics corresponding to this syllabic structure are stored together. Violation of an expectation that is supported by the lexical representation of the word itself leads to more processing difficulties (larger delay in reaction times) than when there is no specific lexical support for the expectation (as is the case for pseudowords).

To conclude, the results of this experiment show that the prosodic mismatch effect is not restricted to the number decision task, but is also visible in auditory lexical decision. It is clear that the participants took the prosodic cues into account, even though these cues were irrelevant for making auditory lexical decisions. Interestingly, the correlational analysis revealed that the prosodic mismatch effect was stronger for words than for pseudowords, suggesting lexical storage of prosodic cues to at least some extent.

General discussion

In this study, we investigated whether uninflected and inflected forms have different prosodic characteristics, and whether such characteristics are functional for the listener in distinguishing these forms, by reducing the ambiguity between them. We found that indeed such acoustic differences exist between uninflected and inflected forms, and that listeners are sensitive to them. When prosodic information mismatches segmental information, participants show a delay in processing (Experiment 1, 2, and 3, number decision, and Experiment 5, auditory lexical decision). We refer to this phenomenon as the *prosodic mismatch effect*. In distinguishing singular forms from the stems of their corresponding plural forms, two sources of non-segmental information in particular play an important part: duration and intonation (Experiment 1, 3, and 5). No cues independent from duration and intonation contribute to the prosodic mismatch effect in singulars and plurals (Experiment 4). The acoustic mismatch effect occurs both in singulars and in plurals (Experiment 2), and in words and pseudowords (Experiment 5). The prosodic differences between uninflected forms and the stems of their corresponding inflected forms reduce the ambiguity between these forms. Our results suggest that these acoustic cues help the perceptual system in determining early in the signal whether an inflected (bisyllabic) or an uninflected (monosyllabic) form is heard.

The existence of the prosodic mismatch effect has important consequences for theories of lexical processing and lexical representation. In classical models of lexical processing, the dominant view has been that all phonetic variation in the speech signal is abstracted away from through acoustic-phonetic analysis, in which the speech signal is translated into a string of discrete phoneme-like units. This abstract string constitutes an intervening representational level through which the speech signal is mapped onto representations in the mental lexicon (Pisoni & Luce, 1987). Since the abstract segmental representation of the singular form would be identical to

that of the stem of the plural form, there is no reason why a delay in processing would occur when there is a mismatch between prosodic and segmental information: After acoustic-phonetic analysis, the processing system no longer has access to prosodic information, neither at the prelexical level, nor at the lexical level. Thus, models of speech perception that propose a strictly phonemic account of lexical access are challenged by the acoustic mismatch effect observed in the present study.

An alternative account of lexical processing and representation, originally proposed as an answer to the inability of the conventional models to deal with phonological variation, abandons the notion of an intervening segmental level (Lahiri & Marslen-Wilson, 1991; Gaskell & Marslen-Wilson, 1996, 1998). Instead, it assumes that the input to the lexical level is featural. It furthermore assumes that there is a single phonological underlying representation for each lexical item, which abstracts away from all surface detail, and which is compatible with all phonologically permissible variants in a given context. The lexical representations in this framework contain only distinctive and marked information. Predictable information is not specified. For instance, in English a word-final /n/ can be realized as /n/, as /m/, or as /ŋ/, depending on the place of articulation of the following segment: *green berry* (/m/), *green glass* (/ŋ/) versus *green dress* (/n/). Hence, the final nasal of *green* is unspecified for place of articulation. In other words, in this framework, phonemic variation is not represented lexically if it is predictable. This suggests that predictable variation that is prosodic in nature is not represented lexically either. If so, it is unclear how the prosodic mismatch effect might arise in this kind of approach.

An approach which can account for the prosodic mismatch effect is that of Johnson (1997). He trained a connectionist (exemplar-based) model on vector quantized speech data, which contained — among other things — information regarding the durations of the segments. Johnson's model correctly anticipated whether the incoming syllable was followed by another (unstressed) syllable or not. Davis et al. (2002) also favor a subsymbolic model that is sensitive to subphonemic properties of the acoustic input.

Our explanation for the occurrence of the prosodic mismatch effect is framed in the exemplar-based or episodic approach of Goldinger (1998), but it can be incorporated in other theoretical approaches as well. We think that in parallel to the processing of the acoustic signal of the stem, an expectation regarding the number of unstressed syllables that will follow is built up based on the

durations of the segments. A delay in processing will occur when this expectation is violated by the segmental material that either does or does not follow the stem. The build-up of an expectation regarding the possible continuation of the signal would be advantageous at several levels.

First, it would provide information regarding the prosodic make-up of the utterance. Salverda et al. (in press) point out that subtle acoustic cues may signal the presence or absence of a prosodic word boundary. They argue that a prosodic representation is computed, based in part on these acoustic cues and in parallel to the segmental encoding. This prosodic representation would contribute to lexical activation by favoring candidates whose boundaries are aligned with the hypothesized prosodic boundary.[1]

Furthermore, the expectation about whether an unstressed syllable is to follow would also provide information regarding the morphological make-up of the incoming speech signal. The prosodic cues signal whether the acoustic signal at hand is that of an unmodified (monosyllabic) stem or that of the same stem but now followed by an unstressed (inflectional or derivational) suffix or by an (unstressed) clitic. We showed that listeners probably determine whether a stem is part of a morphologically simplex form or not, long before the segmental information comes in that signals the presence or absence of a suffix (or clitic).

If it is true that the prosodic mismatch effect arises from the violation of an expectation that is based on the durations of segments, then the question arises how it is possible that listeners are sensitive to these durations, given the enormous variability in the temporal structure of speech. Speech rate varies between speakers, within speakers, and within speakers even within one sentence. Hence, the absolute durations of segments will vary tremendously from utterance token to utterance token. We think that the solution of this riddle lies in the relative durations of the segments in the stem.

<div align="center">PLACE FIGURE 1 APPROXIMATELY HERE</div>

Consider Figure 1, which summarizes the distributions of durations by means of boxplots of the onset, the vowel, and the coda of the monosyllabic stems of the words from Experiment 1 (upper panel) and pseudowords from Experiment 7 (lower panel). The unmodified singulars are denoted by **sg**, and the stems in plurals are denoted by **pl**. The boxes show the interquartile range, the horizontal line in the box denotes the median, and the 'whiskers' extend to the observations

within 1.5 times the interquartile range. Outliers beyond this range are represented by individual circles. Differences in duration that are significant in two-tailed pairwise *t*-tests as well as in two-tailed paired Wilcoxon tests ($p < 0.0001$) are marked with asterisks.

What Figure 1 shows is that there is no reliable difference in duration between the onset of the singular form and the onset of the stem of the corresponding plural form. For the pairs of onsets of existing words, there is a 5 ms difference that fails to reach significance ($t(47) = 1.7, p = 0.10$). For the onset pairs in pseudowords, there is a 6 ms difference in the opposite direction (the onsets of stems in plurals tend to be longer than those of their singulars) that also does not reach significance ($t(46) = -1.5, p = 0.14$).

These small and non-significant differences in duration of the onset contrast with the longer and very significant difference in duration for the vowels (18 ms for the words and 15 ms for the pseudowords). For the codas, the difference in duration is even greater (63 ms for the words and 86 ms for pseudowords, most of which is due to the release noise duration of the final plosive). Considered jointly, this pattern of results suggests that the duration of the onset is a stable anchor point against which the duration of the vowel as well as the duration of the coda can be calibrated. If the durations of vowel and coda compared to that of the onset are relatively long, the incoming speech signal is likely to be a singular. If these durations are relatively short, the likelihood increases that it will be part of a morphological continuation form. In other words, we think that the relative durations of vowel and coda with respect to the onset provide the acoustic information that in our experiments gives rise to the prosodic mismatch effect.

Relative durations, however, differ from word to word. For instance, the relative duration of the vowel with respect to the onset will depend on whether the vowel is phonemically long or short, as well as on the number of segments in the onset. Similarly, the relative length of the coda varies with the number of segments in the coda and in the onset. In addition, specific combinations of segments in the syllable may affect their duration (Waals, 1999). We therefore hypothesize that the relevant information is provided lexically, with a given lexical form, in our experiment a given singular or its plural, having a prototypical distribution of relative segmental durations. In other words, we propose that a lexical entry does not only specify the segments and their order, but also the relative durations of vowel and coda with respect to the onset.

Storage of word-specific prosodic details does not preclude the existence of a general rule or

of an analogical mechanism for building up an expectation of whether an unstressed syllable will follow. In fact, the occurrence of the prosodic mismatch effect for pseudowords (Experiment 5) shows that such a rule or mechanism must exist, as no lexical entries are available for pseudowords. Given an analogical mechanism that generalizes over stored exemplars, the prosodic mismatch effect in pseudowords can be viewed as resulting from implicit knowledge of prosodic structure that emerges from the patterns that are present in the lexicon. This view is supported by the finding that the correlation between prosodic (durational) mismatch and reaction times was stronger for words than for pseudowords, suggesting specific lexical support for the prosody-based expectation regarding the number of syllables to follow for existing words.

The hypothesis of lexical storage of durational information is also supported by the pattern of frequency effects in our experiments, a pattern which strongly suggests that the prosodic information in the stem codetermines which of two representations (singular or plural) becomes most active. For all experiments, we conducted multi-level covariance analyses (Pinheiro & Bates, 2000) in which reaction times were predicted by Duration, Singular Surface Frequency, Plural Surface Frequency, and — where applicable — Durational Difference (between normal and constructed form). We will only discuss the effects of Singular Surface Frequency and of Plural Surface Frequency here.

In the number decision experiments, we observed effects of Singular Surface Frequency in all cases *except* when both the segmental and the prosodic information pointed to a plural form (i.e., in the case of the normal plural forms in Experiment 2). In other words, if any source of information (segmental or prosodic) in the acoustic signal points to a singular form, the singular representation is activated, even when there is a mismatch between the different sources of information in the signal.[2] Plural Surface Frequency, on the other hand, has an effect whenever the prosodic information points to the plural form, irrespective of what form the segmental information points to (i.e., in the case of the constructed singular forms in Experiment 1 and the normal plural forms in Experiment 2).[3] In other words, in a number decision task, the prosodic information in the stem appears to codetermine whether the singular or the plural representation is activated: Prosodic cues to the plural form lead to activation of the plural representation, prosodic cues to the singular form lead to activation of the singular representation.

Note that when, in a number decision experiment, segmental information points to a singu-

lar form whereas prosodic information points to a plural form (i.e., in the case of the constructed singular forms in Experiment 1), we observe competition between the singular and the plural form: Both the singular and the plural representations are activated. In the normal case (i.e., in the case of the normal singular forms in Experiment 1 and in the case of the normal plural forms in Experiment 2), no competition is observed: Only the correct representations are activated. The ambiguity between the singular and the plural form is resolved through the prosodic differences in the stem. This finding solves the functional paradox for competition models of having stored lexical representations for inflected forms in lexical memory. Without prosodic differences between the stem and its stored inflectional continuation form, recognition would be slowed down due to prolonged competition between these forms. Given the prosodic differences documented in this study, the inflected form is a less strong cohort competitor for the uninflected form and vice versa. The existence of a full form representation of the inflected form is therefore no longer highly dysfunctional from the perspective of lexical competition.

In the lexical decision experiment (Experiment 5), we observed a different pattern of frequency effects. There were facilitatory effects of both Singular Surface Frequency and Plural Surface Frequency, for both normal and constructed singular forms. [4] We observed no competition, contrary to in the number decision experiments. Interestingly, for lexical decision, the relevant information is whether the perceived segments form an existing word. As the distinction between the singular and the plural is irrelevant in lexical decision, the support for the singular and plural is pooled: Both the singular and the plural representations support a positive lexical decision.

The hypothesis that durational structure is part of the lexical representations of words is compatible with Goldinger's (1998) episodic (or exemplar-based) theory, according to which experience with spoken word tokens leaves detailed traces of these tokens in memory. It is also compatible with the linguistic distributional evidence brought together by Bybee (2001), evidence which shows that phonologically redundant information is stored in the (mental) lexicon. Furthermore, it is compatible with Pierrehumbert's exemplar-based framework (2002), in which each individual word has an associated probability distribution (exemplar cloud) for each of its segments.

The prosodic mismatch effect documented in this study has important consequences for our understanding of the morphological structure of complex words. The way words are written in languages such as Dutch and English suggests that they consist of stems and affixes that are

strung together as beads on a string. Phonemic transcriptions convey the same impression. Our experiments show that this impression is wrong. Plurals are not just singulars with an additional suffix. The precise acoustic realization of the stem provides crucial information to the listener about the morphological context in which the stem appears.

References

Alegre, M. and Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40:41–61.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.

Baayen, R. H., Burani, C., and Schreuder, R. (1997). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1996*, pages 13–34. Kluwer Academic Publishers, Dordrecht.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 37:94–117.

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R. (in press). Dutch inflectional morphology in spoken- and written-word recognition. In Baayen, R. and Schreuder, R., editors, *Morphological structure in language processing*, pages in press. Mouton de Gruyter, Berlin.

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H., Schreuder, R., De Jong, N. H., and Krott, A. (2002). Dutch inflection: the rules that prove the exception. In Nooteboom, S., Weerman, F., and Wijnen, F., editors, *Storage and Computation in the Language Faculty*, pages 61–92. Kluwer Academic Publishers, Dordrecht.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., and Hyönä, J. (1999). Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language. *Cognition*, 74:B13–B25.

Boersma, P. (1996). *Praat: Doing phonetics by computer.* University of Amsterdam, Amsterdam

Booij, G. E. (1995). *The phonology of Dutch*. Clarendon Press, Oxford.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press, Cambridge.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. LOT, Amsterdam.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28:218–244.

Dressler, W. U., Mayerthaler, W., Panagl, O., and Wurzel, W. U. (1987). *Leitmotifs in Natural Morphology*. Benjamins, Amsterdam.

Fowler, C. A. (1977). *Timing control in speech production*. Indiana University Linguistics Club, Bloomington, Indiana.

Gaskell, M. G. and Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22:144–158.

Gaskell, M. G. and Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24:380–396.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105:251–279.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, 50:101–113.

Lahiri, A. and Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38:245–294.

Lehiste, I. (1972). *Suprasegmentals*. MIT Press, Cambridge, Mass.

Lindblom, B. and Rapp, K. (1973). Some temporal regularities of spoken swedish. *PILUS*, 21:1–59.

Lorch, R. F. and Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16:149–157.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In Altmann, G. T. M., editor, *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, pages 148–172. The MIT Press, Cambridge, Mass.

Marslen-Wilson, W. D., Moss, H. E., and van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22:1376–1392.

McCarthy, J. J. and Prince, A. (1993). Generalized alignment. In Booij, G. E. and van Marle, J., editors, *Yearbook of Morphology*, pages 79–154. Kluwer Academic Publishers, Dordrecht.

McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, 29:225–239.

Nooteboom, S. G. (1972). *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*. University of Utrecht, Utrecht.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52:189–234.

Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology VII*, pages 101–140. Mouton de Gruyter, Berlin.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer, New York.

Pisoni, D. B. and Luce, P. A. (1987). Acoustic-phonetic representation in word recognition. *Cognition*, 25:21–52.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., and Gremmen, F. (1999). How to deal with 'the language-as-fixed-effect-fallacy': Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41:416–426.

Salverda, A., Dahan, D., and McQueen, J. (in press). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*.

Schreuder, R., De Jong, N. H., Krott, A., and Baayen, R. H. (1999). Rules and rote: beyond the linguistic either-or fallacy. *Behavioral and Brain Sciences*, 22:1038–1039.

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

Waals, J. (1999). *An experimental view of the Dutch syllable*. Holland Academic Graphics, The Hague.

Footnotes

1. Salverda et al. (in press) treat the durational difference between the first syllable in *hamster* and *ham* as a result of prosodic lengthening only. We think that both a lengthening and a shortening process play a role. If the durational reduction were indeed solely a consequence of the absence of prosodic lengthening, then the number of syllables following the syllable should not affect the amount of reduction in that syllable, as prosodic lengthening is normally restricted to the final syllable. This is contrary to fact: Nooteboom (1972) showed that the more syllables are added, the greater the durational reduction is in the first syllable.

2. In Experiment 1 (durational plus intonational difference between normal and constructed singular forms), Singular Surface Frequency had a facilitatory effect on reaction times to both the normal singular forms (with segmental and prosodic cues pointing to the singular; $t(1044) = -2.7, p < 0.01$) and the constructed singular forms (with segmental cues pointing to the singular but prosodic cues pointing to the plural; $t(1034) = -3.8, p < 0.001$). The higher the Singular Surface Frequency, the easier it was for participants to give the response 'singular' to both the normal and the constructed singular forms. In Experiment 2 (durational plus intonational difference between normal and constructed plural forms), we observed a facilitatory effect of Singular Surface Frequency ($t(998) = -2.9, p < 0.01$) for the constructed plural forms only (i.e., for the forms that carried the prosodic characteristics of the singular). In Experiments 3 and 4, which have in common that the two types of singular forms do not differ in duration (and in fact both carry the durational characteristics of the singular), we observed facilitatory effects of Singular Surface Frequency for both normal and constructed singular forms (Experiment 3: $t(2269) = -2.8, p < 0.01$; Experiment 4: $t(1699) = -3.4, p < 0.001$). Apparantly, whenever any source of information in the acoustic signal (be it segmental or prosodic information) points to a singular form, the singular representation is activated.

3. In Experiment 1, Plural Surface Frequency had an inhibiting effect on the reaction times to the constructed singular forms only (i.e., to the forms that carried the prosodic characteristics of the plural). The higher the Plural Surface Frequency, the more *difficult* it was for participants to give the response 'singular' to the constructed singular forms. In Experiment 2, we observed a facilitatory effect of Plural Surface Frequency for the normal plural forms ($t(999) = -2.7, p <$

0.01). In Experiments 3 and 4, there were no effects of Plural Surface Frequency, neither for the normal forms nor for the constructed forms (Experiment 3: $t(2269) = 0.7, p = 0.50$; Experiment 4: $t(1699) = 1.4, p = 0.18$). To summarize, whenever prosodic information in the acoustic signal points to the plural, the plural representation is activated, irrespective of what form the segmental information points to.

4. In Experiment 5 (lexical decision), we observed facilitatory effects of both Singular Surface Frequency and Plural Surface Frequency for normal singular forms (Singular Surface Frequency: $t(874) = -3.7, p < 0.001$, one-tailed; Plural Surface Frequency: $t(874) = -2.7, p < 0.01$, one-tailed) and for constructed singular forms (Singular Surface Frequency: $t(864) = -1.6, p = 0.05$, one-tailed; Plural Surface Frequency: $t(864) = -1.8, p < 0.05$, one-tailed). (We applied one-tailed tests as frequency effects are always facilitatory for lexical decision.)

Tables

Table 1: Experiment 1 – Mean duration (in ms) with *SD* and mean average fundamental frequency (in Hz) with *SD* for normal singular forms and constructed singular forms

Type of singular form	Duration	SD Duration	F0	SD F0
Normal	388	73	185	18
Constructed	301	73	192	17

Table 2: Experiment 1 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal singular forms and constructed singular forms

Type of singular form	RT	SD	Error
Normal	335	44	2.0
Constructed	444	36	1.4

Table 3: Experiment 2 – Mean duration (in ms) of the whole word with *SD* and mean average fundamental frequency (in Hz) of the beginning portion of the word with *SD* for normal plural forms and constructed plural forms

Type of plural form	Duration	SD Duration	F0	SD F0
Normal	487	77	190	5
Constructed	516	75	179	7

Table 4: Experiment 2 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal plural forms and constructed plural forms

Type of plural form	RT	SD	Error
Normal	299	50	1.2
Constructed	323	53	1.0

Table 5: Experiment 3 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal singular forms and constructed singular forms when the two types of singular forms only differed in intonational contour

Type of singular form	RT	SD	Error
Normal	333	51	1.7
Constructed	343	47	0.9

Table 6: Experiment 3 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal singular forms and constructed singular forms when no intonational information was carried by the signal

Type of singular form	RT	SD	Error
Normal	357	59	0.3
Constructed	447	58	0.6

Table 7: Experiment 4 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal singular forms and constructed singular forms when no intonational and durational differences were present between both types of singular forms

Type of singular form	RT	SD	Error
Normal	420	52	2.0
Constructed	409	49	1.4

Table 8: Experiment 5 – Mean duration (in ms) with *SD* and mean average fundamental frequency (in Hz) with *SD* for normal singular words and pseudowords and constructed singular words and pseudowords

Type of singular form	Duration	SD	F0	SD F0
Normal word	388	73	185	18
Constructed word	301	73	192	17
Normal pseudoword	451	67	198	14
Constructed pseudoword	357	66	203	13

Table 9: Experiment 5 – Mean response latencies (in ms) measured from word offset with *SD* and Error Percentages for normal and constructed singular word items and for normal and constructed singular pseudoword items

Item type	RT	SD	Error
Normal singular word item	442	79	2.3
Constructed singular word item	531	89	3.3
Normal singular pseudoword item	524	87	2.0
Constructed singular pseudoword item	583	79	4.3

Figure captions

Figure 1. Duration in ms of the onset, the vowel, and the coda in singulars (sg) and in the stems of plurals (pl), for words (upper panel) and pseudowords (lower panel).
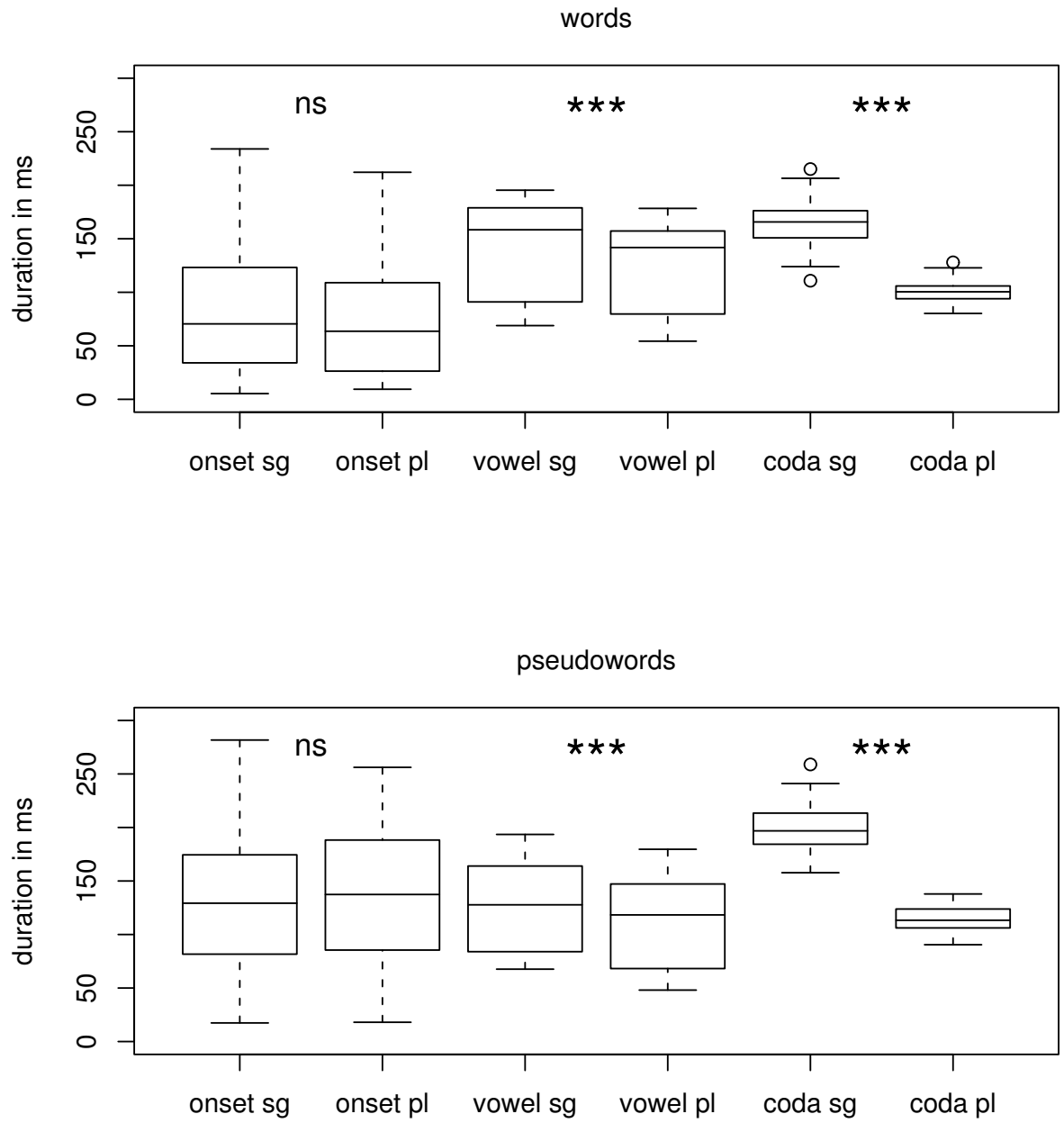
Figure 1.