Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 66 (2015) 65–81

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation

Katrin Schweitzer [a,*], Michael Walsh [a], Sasha Calhoun [b], Hinrich Schütze [a,1], Bernd Möbius [c], Antje Schweitzer [a], Grzegorz Dogil [a]

[a] *Institute for Natural Language Processing (IMS), University of Stuttgart, Pfaffenwaldring 5b, 70569 Stuttgart, Germany*
[b] *School of Linguistics and Applied Language Studies, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand*
[c] *Department of Computational Linguistics and Phonetics, Saarland University, Postfach 15 11 50, 66041 Saarbrücken, Germany*

Received 11 January 2013; received in revised form 19 September 2014; accepted 22 September 2014
Available online 30 September 2014

Abstract

In Germanic languages like English and German, intonation is usually thought to be ‘post-lexical’. That is, it is usually assumed that the choice of intonation contour and the form of the realised contour itself are largely independent of the words used. We present three corpus experiments which show clear evidence of lexical storage of intonation, contrary to these assumptions. Specifically, in each experiment, we show that distributional properties of words affect the prosodic realisation of those words, including accent and boundary placement, and the shape of pitch accents. The first experiment looks at the frequency of occurrence of a given word with a particular pitch accent type and its effect on the shape of accents on that word. We found that the more frequently a word and an accent type appear together, the greater the amplitude of the accent. The second experiment investigates the effect of both the absolute and relative frequency of occurrence of a given word with a particular accent type and their effect on the variability of the shape of these accents. We found that while absolute frequency increases the variability in pitch accent shape, relative frequency reduces it. The final experiment looks at the effect of the relative frequency of a word in its lexical (trigram) context on both variability in its prosodic context and on accent shape variability. We found that both kinds of prosodic variability decrease as the relative frequency of the word in its lexical context increases. We argue that all of these findings are expected within an exemplar approach assuming storage of tonal information with lexical items, and discuss the implications of this for the production and mental representation of intonation.
© 2014 Elsevier B.V. All rights reserved.

Keywords: Intonation; Exemplar Theory; Frequency effects; Prosody

1. Introduction

In Germanic languages like English and German, it is usually assumed that the choice of intonation contour, and the form of the pitch contour itself, are independent of the words used (save for pitch peak alignment effects caused by the syllabic form of the words, etc.). Our research, conducted within the framework of Exemplar Theory, provides strong evidence for frequency-based storage of intonation with words – suggesting that the

[*] Corresponding author. Tel.: +49 711 685 8 4589.
E-mail addresses: katrin.schweitzer@ims.uni-stuttgart.de (K. Schweitzer), michael.walsh@ims.uni-stuttgart.de (M. Walsh), sasha.calhoun@vuw.ac.nz (S. Calhoun), inquiries@cislmu.org (H. Schütze), moebius@coli.uni-saarland.de (B. Möbius), antje.schweitzer@ims.uni-stuttgart.de (A. Schweitzer), dogil@ims.uni-stuttgart.de (G. Dogil).

[1] Current address: Center for Information and Language Processing, University of Munich (LMU), Oettingenstr 67, 80538 München, Germany.

http://dx.doi.org/10.1016/j.specom.2014.09.006
0167-6393/© 2014 Elsevier B.V. All rights reserved.

choice of intonation contour, and its form, are not independent of the words used.

Below, we first describe Exemplar Theory in Section 1.1, followed by a discussion of its relationship to autosegmental-metrical models of intonation (Section 1.2), and how it can be employed to investigate the possibility of lexicalised storage of intonation (Section 1.3). In Section 2 we set out the findings from three corpus studies examining intonation parameters with respect to possible lexical frequency effects. We examined if the frequency with which lexical items and pitch accents occur together is related to the detailed acoustic realisation of the pitch accent tokens. Moreover, we examined if the relative frequency of a word in its trigram context is related to how it is realised on the tonal level. If distributional properties of the lexical level have an influence on intonation, this is consistent with storage of intonation features with the lexical items. The results of these studies are discussed in Section 3.

1.1. Exemplar Theory

Exemplar Theory was introduced in psychology to model perception and categorisation (Kruschke, 1992; Medin and Schaffer, 1978; Nosofsky, 1986; Nosofsky et al., 1992). In recent years, it has also been applied to speech perception (Goldinger, 1997; Johnson, 1997) and production (Pierrehumbert, 2001; Wade et al., 2010; Walsh et al., 2010). The key idea of Exemplar Theory in language and speech is that linguistic stimuli are stored as highly detailed episodes which are then employed in speech production and perception. Exemplars are assumed to carry information about, e.g., the detailed acoustics, the speaking situation, the context when that exemplar was used, etc.

New exemplars are categorised by similarity to exemplars stored in memory. Stored exemplars are accessed as production targets. The exemplar memory is constantly updated, and is highly sensitive to frequency and recency effects. Frequent units are represented by many exemplars, infrequent units are represented by few exemplars. The variance that occurs within a category is implicitly encoded: all the realisations of a given unit are stored and will influence new productions. Frequency effects have been shown to occur in a variety of linguistic domains, and phonetic research has documented various phonetic parameters, articulatory as well as acoustic, which are influenced by frequency of occurrence (Bybee and Scheibman, 1999; Carreiras and Perea, 2004; Cholin et al., 2006; Jurafsky et al., 2001; Losiewicz, 1992; Pluymaekers et al., 2005).

Effects of frequency on stored exemplars are well documented on the lexical level, where frequency influences segmental parameters. For instance, Losiewicz (1992) found a dependency between the duration of the past tense morpheme *-ed* in segmentally similar (i.e. rhyming) verbs and the lexical frequency of the verbs under investigation, with the morpheme being shorter in high-frequency verbs. Pluymaekers et al. (2005) found effects of word frequency on the duration of three Dutch affixes in a corpus study that controlled for effects of speech rate, position of the word in the utterance, disfluencies, and phonetic contexts. The affixes were shorter (either in affix duration or in segment duration or both) in high frequency words. Bybee and Scheibman (1999) carried out a study demonstrating that the degree of the phonetic reduction of the word *don't* in conversational speech is greater in frequent collocations. A similar relationship was found by Bybee (2000) who showed that the rate of *t*/*d*-deletion is related to the lexical frequency of the word in question. Further, Jurafsky et al. (2001) found the relative frequency of a word in its lexical context to be positively correlated with phonetic reduction. For a detailed overview of results on phonetic reduction see Ernestus (2014).

There is also evidence that frequent units display less variation. For instance, during language acquisition, the variability of a phonetic category decreases (Lee et al., 1999). From an exemplar-theoretic perspective, such a finding can be accounted for by *entrenchment* – decreasing variability with increased production practice (implicitly captured by an increased number of exemplars). Pierrehumbert (2001) illustrates the phenomenon with the example of a child learning a stringed instrument: whereas in the beginning the notes will be highly variable, after years of practise (having produced and stored a large number of exemplars) there will be considerably less variance.

Given the frequency effects discussed above, the exemplar-theoretic assumption that linguistic events are stored as single instances, rich in detail, seems plausible. Consequently, fundamental frequency could also form part of stored representations. On this assumption, in this article we adopt an exemplar-theoretic approach to investigating how distributional properties of words are related to their tonal realisation.

1.2. Autosegmental-metrical theory of intonation and exemplar-theoretic effects

The most widespread models of intonation (e.g. Silverman, 1992, Baumann et al., 2001 often subsumed under the term *autosegmental-metrical (AM) models*, predominantly building on Pierrehumbert, 1980), are silent on any effects of frequency of occurrence. As the name suggests, these models assume that intonation is assigned autonomously from the segmental level according to top-down information such as syntactic, semantic or pragmatic information (e.g. Ladd, 2008). That is, on the basis of such information, a sequence of *tones* is chosen from a discrete set of phonological categories for that language variety. These tones are then associated with the words in the utterance according to their metrical structure. The separation of the lexical and the tonal level in English is explicitly stated by Pierrehumbert (2000, p. 20): "...pitch accents are not underlying properties of words. Instead, they are independent pragmatic morphemes which are co-produced with words."

AM models assume that the realisation of the pitch contour is then determined by phonetic rules which refer solely to the sequence of tones and the metrical organisation of the utterance, e.g. syllables in the word, location of word stress (as well as a module specifying the pitch range of the phrase). Therefore, the phonetic realisation of pitch accents and boundary tones is rule-based, and should not be influenced by the particular words chosen.

Consequently, the assumptions of AM theory are at odds with the idea of storing acoustic detail together with the lexical item, as suggested by exemplar models. If pitch-accenting is assumed to be post-lexical and solely rule-based, the frequency of occurrence interactions between intonation and the lexical level presented below are hard to explain. There is no inherent reason within the theory why the choice of tonal contour and its phonetic realisation should be affected by the frequency of occurrence of the word.

1.3. Lexicalised storage of intonation

If the basic principle of exemplar models, i.e. storage of rich (phonetic) detail, holds for fundamental frequency, then the stored exemplar representations contain pitch contour information, that is, pitch information can be stored lexically.

However, current exemplar models do not take account of intonation (e.g. Johnson, 1997; Pierrehumbert, 2001; Wade et al., 2010; Walsh et al., 2010). Further, there has been very little research on frequency effects on prosody. However, there is evidence that the acquisition of the prosodic word is frequency driven (Vigário et al., 2006), that word stress assignment can be inferred by instance-based learning (Daelemans et al., 1994), and that distributional properties of syllables influence the predictability of their durations (Schweitzer and Möbius, 2004; Walsh et al., 2007). For tonal parameters, there is very little research investigating frequency effects, with the exception of Braun et al. (2006) who showed that random tonal contours gravitate towards frequent contours in an iterative mimicry study. There is also very little research that explicitly tackles the question whether intonation can be stored lexically. Goldinger (1997) mentions a pilot study in which speakers in a shadowing experiment adapted their pitch to the pitch of the stimuli. Such a result can be interpreted as evidence for storage of F_0. Moreover, Calhoun and Schweitzer (2012) present a corpus study paired with a perception experiment that indicates lexical storage of intonation.

Besides these two studies which come from an exemplar-theoretic angle, there is other work that indicates storage of sentential intonation. These studies are firstly situated in the domain of psycholinguistics, where several experiments demonstrate that the familiarity or frequency of prosodic parameters influence speech processing, perception and production (Braun et al., 2006; Braun and Johnson, 2011; Braun et al., 2011; Mandel et al., 1994; Van Lancker and Canter, 1981; Van Lancker et al., 1981). A second research area which provides evidence for lexicalised storage of intonation is the area of machine learning, where various studies showed that word identity helps in predicting pitch accent location (Brenier et al., 2006; Nenkova et al., 2007; Pan and Hirschberg, 2000; Pan and McKeown, 1999), and where instance-based learning of prosody outperforms other types of learning (Marsi et al., 2003).

We can see, then, that while there is a range of circumstantial evidence for frequency-based storage of intonation with lexical items, this has not been tackled directly in most previous research. The work presented below attempts to do this by investigating the relationship between frequency of occurrence and intonation.

2. Corpus analyses

If exemplar-theoretic assumptions hold, and intonation can indeed be stored with segmental information, i.e. with the word itself or with sequences of several words, then frequency effects on tonal parameters would be expected; just as has been found in the segmental domain.

This prediction is explicitly targeted in the experiments presented below. They test for potential dependencies between both the prosodic pattern used and the pitch accent shape and frequency of occurrence. Both the frequency of the combined type of word and tonal event, as well as the influence of pure lexical frequency on pitch accent shape, are examined.

More specifically, experiment 1 demonstrates how pitch accent realisation is influenced by the frequency of the word + accent pair, i.e. the combination of a word and the accent it occurs with (cf. Schweitzer et al., 2010a). Experiment 2 then looks at the relative frequency of such pairs in production. High relative frequency implies that of all the instances of that word the majority occurs with the respective accent. From an exemplar-theoretic viewpoint this entails that the majority of available production targets carries the same accent, which could lead to differences in variability. The experiment shows that indeed the production of pitch accent tokens can be influenced by the relative frequency of the pairs (cf. Schweitzer et al., 2010b), in addition to their absolute frequency. Finally, experiment 3 investigates the relationship between the relative frequency of the middle word in a given sequence of three words and the prosodic variability of the word sequence, demonstrating that word sequences that occur together relatively often display less prosodic variability (cf. Schweitzer et al., 2011).

An overview of the experiments is given in Table 1. All experiments relate a tonal parameter to a frequency measure. To get a comprehensive picture of the nature and validity of frequency of occurrence effects at various levels, different units and different ways to calculate frequency are examined. In order to get as much data as possible for each experiment, we used one corpus for the first two experiments, and a different corpus for the third. For the first two experiments, we needed to use a corpus with tonal

Table 1
Overview of the experiments. Various tonal parameters, frequency measures, languages and speaking styles are examined to achieve a comprehensive picture. (PA = Pitch accent).

	Experiment 1	Experiment 2	Experiment 3
Unit	PA + word	PA + word	word + context
Frequency measure	Absolute	Absolute and relative	Relative
Tonal parameter	Pitch Accent (PA) shape	PA shape variability	PA shape variability prosodic context variability
Language	German	German	American English
Genre	Radio	Radio	Spontaneous

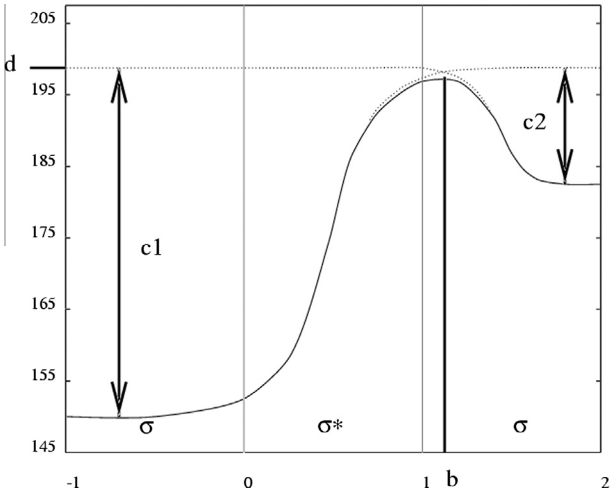

Fig. 1. The PaIntE model function in a three-syllable window, normalised for syllable length. The starred syllable σ^* is the one bearing the pitch accent. Four of the six free parameters of the PaIntE function are marked in the figure: while $c1$ and $c2$ denote the amplitude of the accent's rise and fall (in Hz), parameters d and b define the temporal alignment (b) and the height (d) of the peak. The two additional parameters $a1$ and $a2$ are not displayed: they give the amplitude-normalised steepness of the rising and falling slope, respectively. (Figure from Möhler, 2001, p.2)

pitch accent type annotations, which restricted the choice of corpora substantially. For experiment 3, however, this was not necessary, so we used the largest available suitable corpus. This had the added advantage that we could ensure that the effects found were not language or speaking style specific: German and American English were examined, in news broadcasts as well as spontaneous speech.

To determine pitch accent shape, a parametric intonation model ("PaIntE", see Möhler, 2001) was employed. The remainder of this section first provides a description of the PaIntE model and then describes our three corpus experiments.

2.1. Parametrisation of pitch accent shape

The PaIntE-model ("Parametric Representation of Intonation Events", cf. Möhler and Conkie, 1998; Möhler, 2001) is a data-based approach to intonation modelling, originally implemented for F_0 generation in speech synthesis. The model approximates stretches of F_0 and interpolates between them. The approximation is implemented as a (linguistically motivated) mathematical function with six free parameters defining the tonal contour within an analysis window spanning the syllable marked with a tonal event and potentially the neighbouring syllables. The function is composed of two sigmoids which create a curve with a peak.

An example of a PaIntE generated contour is given in Fig. 1. Here, the function defines a curve that is shaped like a rising accent – like L*H in the GToBI(S) taxonomy (Mayer, 1995). The starred syllable is the accented one. The figure illustrates the linguistic interpretation of the function's parameters: parameter b marks the alignment of the highest point in the curve, i.e. the alignment of the accent's peak, with the syllables. Parameter d determines the height of this peak. Parameters $c1$ and $c2$ model the amplitude of the rising and the falling sigmoid, respectively. Linguistically, they give the range of the accent. Parameters $a1$ and $a2$ estimate the steepness of the rise and the fall, respectively. They are normalised for the amplitude of the respective slope, i.e. the actual values of the gradients are divided by $c1$ or $c2$, respectively (Möhler, 2001).

It is important to note that the PaIntE model allows for different approximation methods. The default case is to model the pitch accent using two sigmoids. However, if the original F_0 is modelled better with only one sigmoid the model tries to approximate the accent using only one (cf. Schweitzer, 2010, for more detail). In such cases, the c parameter for the unused sigmoid ($c1$ or $c2$) is assigned 0, and the respective a parameter for the unused sigmoid ($a1$ or $a2$) is set to -1. Note that those two values differ in whether they are meaningful or not: while it is reasonable to say a non-existing slope has an amplitude of 0 (parameter c), the assignment of -1 for the gradient of the slope does not result in an interpretable value. This has some consequences for the methodology employed in experiments 2 and 3 (see below).

The PaIntE parameters were used to measure changes in pitch accent shape in the following experiments.

2.2. Experiment 1: Absolute frequency of pitch accent + word pair

The first experiment was designed to investigate if and how the pitch contour can be influenced by the word on which it occurs. The frequency with which linguistic units occur has been shown to influence phonetic detail (Bybee and Scheibman, 1999; Carreiras and Perea, 2004; Cholin et al., 2006; Jurafsky et al., 2001; Losiewicz, 1992). Regarding pitch, in previous work (Schweitzer et al., 2010a) we found that the frequency with which a specific word occurs together with a given pitch accent, is related to properties of the pitch accent's realisation: specifically, we found increased accent ranges with increasing frequency of a word + accent pair. In the experiment presented here we analyse a larger dataset (more than 5 h compared to under 3 h in Schweitzer et al., 2010a) with respect to such a correlation between word + accent pair frequency and pitch accent ranges.

2.2.1. Data

The experiment was carried out on the prosodically annotated part of the DIRNDL-Corpus (Eckart et al., 2012) consisting of 55 German radio news broadcasts (5 h and 16 min, 5 male and 4 female speakers). The corpus is annotated for pitch accents and boundary tones according to the GToBI(S) guidelines (Mayer, 1995). It comprises 7817 L*H and 6118 H*L accents. For each word type, the frequency of the combination of this type with an accent type (a *word + accent pair*, e.g. "Berlin + H*L") was calculated. The frequency of word + H*L pairs ranged from 1 to 46, the frequency of word + L*H pairs from 1 to 52. For each pitch accent in the corpus, the PaIntE parameters were calculated.

Then we extracted two datasets, one comprising H*L pitch accents, and one comprising L*H accents. Tokens with outlying values for any of the PaIntE dimensions were removed. Outliers were defined as tokens that fell outside the whiskers in a boxplot, i.e. they were more than 1.5 times the interquartile range (IQR) away from the quartiles. Thus 6588 L*H tokens and 4663 H*L tokens remained in the analysis.

Since the data processing preceding the PaIntE parametrisation includes several steps (manual labelling, F_0 approximation and F_0 smoothing) and since each step increases the number of potential errors, the pitch accent tokens were tested with a very conservative methodology for their plausibility before they were included in the analysis. Only pitch accent tokens that were clear examples of either L*H or H*L were extracted. Clear examples were defined as meeting one of the following criteria:

1. If both of the function's sigmoids were used to model the accent – hence F_0 was approximated according to the two-sigmoid (standard) case – then, for H*L, the amplitude of the fall had to be greater than the amplitude of the rise. Similarly, for L*H, the amplitude of the rising sigmoid had to be greater than the amplitude of the falling one.
2. If only one sigmoid was used, it had to be the falling sigmoid for H*L and the rising sigmoid for L*H.

These plausibility checks restricted the datasets to 2587 tokens (H*L dataset) and 5378 tokens (L*H dataset). The ranges of the frequencies of the word + accent types were unaffected (H*L [1–46], L*H [1–52]). In order to inspect the excluded pitch accent tokens we used WEKA's (Hall et al., 2009) simple *k*-means algorithm to cluster them into four groups. To visually compare the excluded accents and the included ones, we plotted the centroid of the excluded clusters for each accent type along with the centroid for the included accents. Fig. 2 shows the resulting plots for H*L and L*H. As can be seen, the cluster centroids of the included data correspond to the canonical pitch accent types in GToBI(S) (Mayer, 1995): H*L is defined as a peak in the accented syllable followed by a low target in the post-accented syllable. The solid black line in the left graph illustrates that the included accents are plausible examples of H*L. Analogously, L*H is defined as a low target in the accented syllable followed by a rise in the post-accented one. The solid black line in the graph on the right-hand side shows that the included L*H accents are plausible L*H examples. (Note that the PaIntE function does not model valleys in the contour, but only peaks, i.e. we cannot expect to see a valley in the accented syllable, but only the contour coming from the bottom part of the register.) The cluster centroids of the excluded accents in both graphs (dashed/dotted lines), do not correspond to the definition of the respective accent type.

It is important to note that the source of the implausible pitch accents shape could be labelling errors, smoothing errors or cases where for example due to creaky voice, the fundamental frequency contour could not be derived correctly. Also, these could be cases where the PaIntE approximation failed. Investigation of these excluded tokens will be the subject of future work. For the present studies, which rely on an accurate description of the pitch accent shape, we adhered to the above-described, conservative way of data cleaning, even though it constituted a substantial data reduction in the case of H*L accents.

2.2.2. Methodology

The two pitch accent sets were analysed with respect to whether the accent range of the pitch accent tokens increased with increasing frequency of the word + accent pair (Schweitzer et al., 2010a). The accent range is captured by the PaIntE parameter *c1* for L*H accents and by PaIntE parameter *c*2 for H*L accents. We will use the term *accent range* in the following instead of the two specific parameters in order to have a single term for both datasets. We used the *lme4* package (Bates et al., 2013) in R (R Core Team, 2013) to fit linear mixed effect models with accent range as the dependent variable. The frequency of word + accent pairs was tested as a fixed effect. The frequency values were logged and centred, in order to alleviate the problems associated with Zipfian distributions. Several additional linguistic factors which are known to influence accent shape (Jilka and Möbius, 2007; Mücke et al., 2006; Van Santen and Möbius, 2000) were tested as

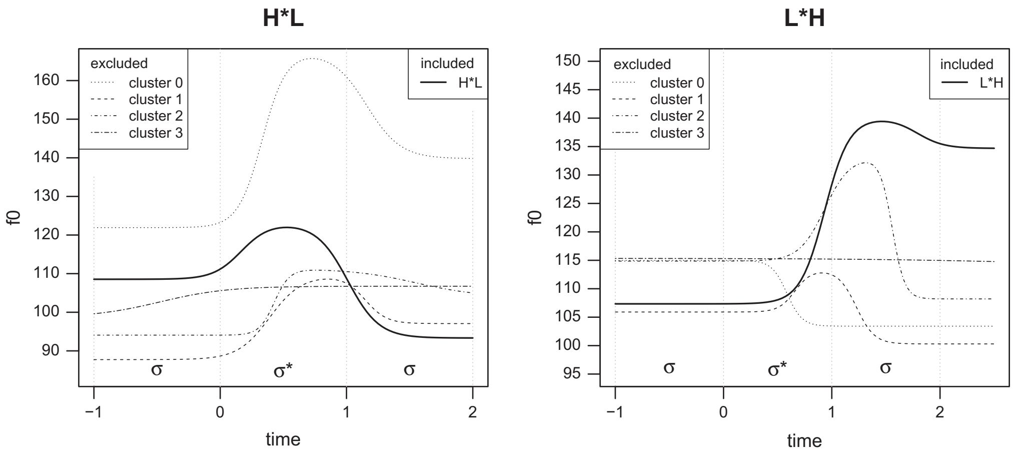

Fig. 2. Plausibility checks. The plots compare the excluded data to the included data. The solid black line in each plot represents the cluster centroid of the pitch accent tokens that were considered to be clear examples of the respective pitch accent type. The dotted and dashed lines represent the cluster centroids of the excluded pitch accent tokens. Token numbers in the excluded L*H clusters are as follows: 324 instances (43%) in cluster 0, 107 (14%) in cluster 1, 196 (26%) in cluster 2, and 118 (16%) in cluster 3. For H*L, cluster 0 comprised 462 instances (27%), cluster 1: 638 (37%), cluster 2: 402 (24%) and cluster 3: 202 (12%).

fixed effects. Firstly, we looked at the number of accents to the next intonation phrase boundary, as a measure of the distance to the end of the tonal phrase whereby the difference between pre-nuclear and nuclear accents can also be captured. In addition, two aspects of syllable structure were taken into account, viz. coda and onset size in terms of number of segments, as well as the Van Santen/Hirschberg-classifications (Van Santen and Hirschberg, 1994) for coda and onset, as determined by the German extension (Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2010) to the Festival speech synthesis system (Black, 1997). The Van Santen/Hirschberg-classification distinguishes onsets and codas with regard to their segmental content, i.e. whether they consist of unvoiced segments only ($-V$), of sonorants only ($+S$), or of voiced segments not including sonorants ($+V$–S). Additionally, we included random intercepts and slopes for word and speaker.

For each data set, we determined the best fitting linear mixed model by carrying out model comparisons using likelihood ratio tests (in a similar fashion to Baayen, 2008; Baayen et al., 2008; Winter, 2013). Starting with the best model comprising only random factors, we added each fixed factor separately, retaining the new model only if the added factor significantly improved it. Subsequently, we tested in the same way for random slopes.

We considered a new model to be better than its predecessor if the improvement was significant ($p(\chi^2) < 0.05$) and if the AIC value (Akaike's information criterion, see Akaike, 1973) was at least 2 points smaller (Burnham and Anderson, 2002).[2]

The p-values reported for each factor were obtained by comparing the winning models to the models without the factor under investigation.

2.2.3. Results

Linear mixed models assess the relationship between the fixed factors and the dependent variable by finding an equation that predicts the value of the dependent variable as a linear combination of the fixed factors plus some intercept, plus possible random effects. The coefficients in this linear combination are estimated from the data. We can then interpret the coefficients for the fixed factors as indicating the expected effect of the factor in question, and the intercept as a kind of "default value" that is expected when all fixed factors are zero (for linear factors, or the 'default' value for categorical factors).

2.2.3.1. H*L accents.

For the H*L dataset, the model term for the linear mixed model fitting the data best is given in Eq. (1). The model incorporates the (logged and centred) frequency of word + H*L pairs *word.accent.freq* as a fixed effect. None of the other linguistic factors tested as fixed effects improved the model significantly. As random effects, the model has intercepts for *word* and *speaker*, represented in the model equation below as $(1|word)$ and $(1|speaker)$, respectively. Additionally, the model had by-speaker and by-word random slopes for *distance.to.phrase.end*, represented by the terms $(0 + distance.to.phrase.end|speaker)$ and $(0 + distance.to.phrase.end|word)$.

$$\begin{aligned} accent.range \sim\ & word.accent.freq + (1|word) + (1|speaker) \\ & + (0 + distance.to.phrase.end|speaker) \\ & + (0 + distance.to.phrase.end|word) \qquad (1) \end{aligned}$$

[2] If the p-value and the AIC value returned contradicting results when adding a factor, we retained the simpler model, but once the final winning model was established, we tested whether this factor could augment the winning model.

An overview of the fixed effects in the model described in Eq. (1) is given in Table 2. Given this model, we expect the intercept value of 31.73 Hz for cases where *word.accent.freq* is zero. The parameter of interest, the frequency with which a word type occurred with an H*L in the data, affects the range of the fall of the H*L accent significantly ($\chi^2 = 4.6243, p < 0.05$ compared to the respective null model). The estimated coefficient is 0.89 Hz ± 0.42 (SE, standard error), i.e. the range is increased by approx. 0.89 Hz for each unit increase in logged frequency of occurrence, i.e. for each multiplication by $e \approx 2.7182$ in the unlogged frequencies. This would for instance correspond to an increase of 3.47 Hz for a word + accent pair with frequency 50 in our dataset compared to one with frequency 1. This increase might sound small at first, however, an increase in accent ranges due to distributional properties must be expected to be very subtle.

*2.2.3.2. L*H accents.* For the L*H dataset, the model term for the linear mixed model fitting the data best is given in (2).

$$\begin{aligned} accent.range \sim\ & word.accent.freq \\ & + distance.to.phrase.end + coda.type \\ & + (1|word) + (1|speaker) \qquad (2) \end{aligned}$$

The model incorporates the (logged and centred) frequency of the word + L*H pair (*word.accent.freq*), the *distance.to.phrase.end*, and the *coda.type* as fixed effects, and random intercepts for *speaker* and *word*. An overview of the model's fixed effects is given in Table 3. The correlation of the fixed effects was very low ($r < 0.1$ as computed by R's *lme4* package, cf. Bates et al., 2013).

The intercept of the model is the value predicted for $word.accent.freq = 0$, $distance.to.phrase.end = 0$, and $coda.type = -V$ (unvoiced). Note that such a setting does not exist in the data, because the current accent is always included when counting accents before the next phrase boundary, i.e. *distance.to.phrase.end* is never below 1 for pitch accented words. The predicted value of the intercept is 33.816 Hz.

The parameter of interest, the frequency of word + accent affected the accent range significantly ($\chi^2 = 10.23, p < 0.005$ compared to the respective null model). The effect is a positive effect, i.e. we expect greater accent ranges for more frequent word + accent pairs. Each unit increase in logged frequency of occurrence (i.e. each multiplication by $e \approx 2.7182$ in the unlogged frequencies) is predicted to increase the range by about 1.08 Hz ± 0.34 SE. This would for instance correspond to an increase of 4.24 Hz for a word + accent pair with frequency 50 compared to one with frequency 1.

Table 2
Estimated coefficients, standard errors (SE), and *t*-values for the linear mixed model predicting accent range for H*L dataset in experiment 1.

	Estimate	SE	*t*-Value
(Intercept)	31.73	3.41	9.32
word.accent.freq	0.89	0.42	2.15

Table 3
Estimated coefficients, SE, and *t*-Value for the linear mixed model predicting accent range for the L*H dataset in experiment 1.

	Estimate	SE	*t*-Value
(Intercept)	33.82	3.18	10.65
word.accent.freq	1.08	0.34	3.21
distance.to.phrase.end	−1.71	0.27	−6.44
coda.type + S	1.91	0.76	2.49
coda.type + V–S	−0.64	5.36	−0.12

Two of the factors tested to control for other known effects on accent shape proved significant. The distance of the accent to the end of the phrase significantly affected accent ranges ($\chi^2 = 41.33, p \ll 0.0001$ compared to the respective null model): As the number of intermittent accents between an accent and the next phrase boundary increases, the predicted range of the accent in question significantly decreases (resulting in a negative coefficient for *distance.to.phrase.end* in the model). In other words, at the beginning of a phrase (where a greater number of accents occur between the accent under investigation and the phrase boundary), accent ranges are smaller, and towards the end of the intonation phrase, accents tend to have higher ranges, possibly reflecting increased acoustic prominence of nuclear accents, which are usually phrase-final. Coda types also significantly affected the accent range ($\chi^2 = 6.3083, p < 0.05$ compared to the respective null model) with sonorant codas having greater ranges. Voiced codas without a sonorant did not differ significantly from voiceless ones ($t = -0.12$, corresponding to $p \approx 0.9$, cf. Table 3).

2.2.3.3. Summary. The results from this experiment demonstrate how pitch is subtly but significantly influenced by the frequency with which a particular pitch accent occurs with a particular word. Specifically the frequency of the combination of pitch accent and word was shown to be significantly related to the accent range, i.e. the amplitude of the rise for rising accents (L*H) and the amplitude of the fall for falling accents (H*L). This finding is in keeping with our earlier work (Schweitzer et al., 2010a), however here we employ a more extensive dataset and a more rigorous methodology.

These effects demonstrate that the word and the tonal level are intertwined and that they are subject to frequency effects. Section 3 elaborates on what implications this result has for models of intonation.

2.3. Experiment 2: Relative frequency of pitch accent + word pair

Experiment 1 showed that pitch accent shape is sensitive to the absolute frequency of pitch word + accent pairs. The second experiment sets out to further investigate how pitch accents are influenced by distributional characteristics of the words on which they are realised, this time looking at *relative* frequency. Segmental parameters have been shown to be related to the relative frequency of a word given its context: the greater the relative frequency, the higher the degree of phonetic reduction (Jurafsky et al., 2001). Phonetic reduction is also related to the relative frequency of a word given the number of phonetically similar words (Wright, 1997), a factor which is also influential on the intelligibility of stimuli (Luce, 1986). With regard to suprasegmental features, Nenkova et al. (2007) showed that the accent ratio, which is derived from the relative frequency with which a word type occurs with an accent in a corpus, is highly predictive of accent placement. In a previous experiment using American English radio broadcast news (Schweitzer et al., 2010b), we showed that the relative frequency with which a word and a given pitch accent type occur together (given all pitch accented instances of the word) influences how variable the realisation of those pitch accent tokens are: the higher the relative frequency, the lower the pitch accent realisation variability. Here, we aimed to replicate this result using a larger corpus of German. In the present experiment, for each combined type of pitch accent and word, we calculated a value which reflects the variability in the realisation of the pitch accent tokens and related this to the relative frequency with which that accent occurs on that word, while controlling for absolute frequency.

2.3.1. Data

For the second experiment, we used the same two datasets as in experiment 1 (cf. Section 2.2.1).

2.3.2. Methodology

The two pitch accent datasets were analysed with respect to whether the variability of the pitch accents changed with the relative frequency of a word + accent pair. To this end, we calculated pitch accent variability amongst all realisations of each word + accent pair and related this value to the relative frequency of word + accent pair.

2.3.2.1. Calculation of pitch accent variability. To measure the variability amongst tokens of word + accent pairs, we used the six PaIntE parameters extracted for each pitch accent token. To normalise the values, the PaIntE parameters were *z*-scored for each speaker and accent type separately, i.e. each PaIntE parameter was represented by a *z*-scored value showing how many standard deviations the raw value was away from the mean value of that parameter for a given speaker and a given accent.

Each pitch accent was then represented as a vector of *z*-scored PaIntE values.

To measure variability, we calculated for each accent + -word pair type which occurred at least twice in the processed datasets, the Euclidean distance to every other pair of the same type. This was calculated according to Eq. (3), where $d(x,y)$ is the Euclidean distance between the vectors x and y.

$$d(x,y) = \sqrt{\sum\nolimits_{dim \in Painte} (x_{dim} - y_{dim})^2},\quad PaIntE = \{a1, a2, b, c1, c2, d\} \tag{3}$$

Then, we calculated the average of these comparisons. That is, for a token that occurred 10 times in the dataset (for instance, “Porsche + L*H”), each token was compared to the remaining 9 tokens and the average of these distances was calculated. The average distance of a pair from all the other instances of the same pair is the variability measure for this token: the smaller the distance to the other tokens of the same type, the greater the similarity. Analogously, a larger value in average Euclidean distance indicates greater variability.

This was straight-forward for the accents which were parametrised using two sigmoids. However, we needed to adapt the method for those cases where the PaIntE model employed only one sigmoid to parametrise one or both accents being compared. Recall that the *a*-values are not meaningful in those cases where accents were modelled using only one sigmoid (cf. Section 2.1). Therefore, we adapted the calculation of Euclidean distance so that a comparison between a two-sigmoid-accent and a one-sigmoid-accent resulted in a categorical distinction on the *a*-dimension – to model the difference between an existing and a non-existing parameter (see Fig. 3 for an illustration of the possible comparison permutations), which is a categorical distinction and should not be measured on a continuous scale. This was achieved by setting the distance between the *a* values of a one-sigmoid-accent and a two-sigmoid-accent to a large constant value. The constant was chosen so that it was higher than the largest difference between a-parameters that occurred in the data set.[3]

Note that in the *z*-scoring preceding the variability calculation the meaningless *a* values were not included.

2.3.2.2. Statistical testing. The H*L and the L*H datasets were again analysed using the *lme4* package in R (Bates et al., 2013). Linear mixed effects models were fitted with variability as the dependent variable. The relative frequency of the word + accent pair was tested as a fixed effect, as was the absolute frequency. Both frequency values were logged and centred. Further, as a control for other factors known to affect accent shape, we again tested the distance to the end of the tonal phrase and the coda type which had been shown to significantly influence the shape of L*H accents in experiment 1. The word was incorporated as a random effect. Note that the data was already normalised for speaker-specific differences because the PaIntE values were *z*-scored (cf. Section 2.3.2.1). For this reason, we did not include speaker as a random effect.

[3] However, it has to be noted that choosing a middle range distance value for the constant, the results remain the same, so changes in variability are not only due to the one-sigmoid cases.

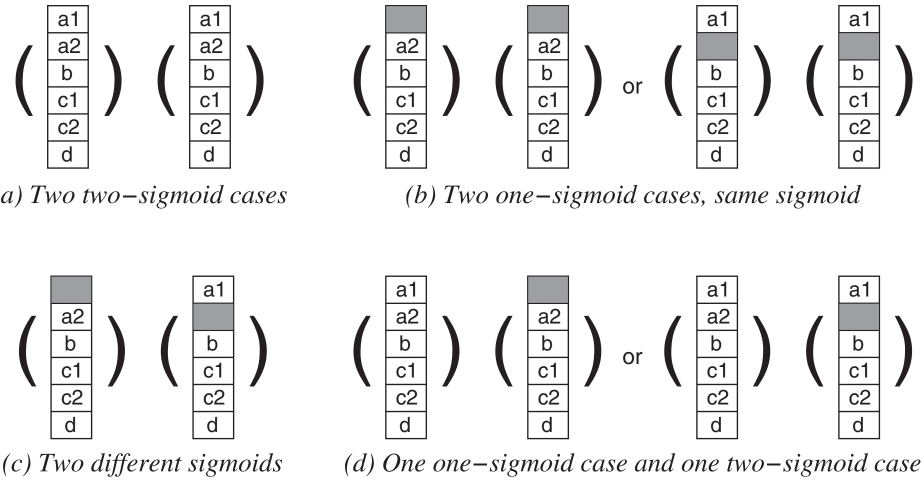
a1
a2
b
c1
c2
d
or
(a) Two two−sigmoid cases
(b) Two one−sigmoid cases, same sigmoid
(c) Two different sigmoids
(d) One one−sigmoid case and one two−sigmoid case

Fig. 3. Possible combinations when calculating the Euclidean distance between two accent tokens (experiments 2 and 3). In the PaIntE approximation, each accent token can either be approximated by two sigmoids, by a falling sigmoid, or by a rising sigmoid. In the two-sigmoid case, all PaIntE values are assigned a meaningful value, in the one-sigmoid case, one of the *a* dimensions is assigned a dummy value (grey box). The calculation of the Euclidean distance between the two respective tokens varies for the different combinations: in case (a) Euclidean distance was determined as usual, in case (b), the non-meaningful *a* dimension was ignored, and in cases (c) and (d), the distance on the respective *a* dimension was set to a large constant. Note that case (c) cannot occur in experiment 2 because variability is calculated among accents of the same type and the plausibility tests exclude cases where the accent was parametrised with the wrong sigmoid. However, it can occur in experiment 3 where pitch accent type is not examined.

Analogously to experiment 1, we determined the best fitting model using likelihood ratio tests (Baayen, 2008; Baayen et al., 2008; Winter, 2013). Starting with the simplest model, we tested each factor to see if its inclusion made the model a significantly better fit to the data. When comparing two models, we assumed a model was better than its competing simpler model if the improvement was significant ($p < 0.05$) and if the AIC value was at least 2 points smaller; p-values reported for each factor were obtained by comparing the null models to the models with the respective factor.

2.3.3. Results

*2.3.3.1. H*L accents.* For the H*L dataset, the model term for the linear mixed effects model fitting the data best is given in Eq. (4). The model incorporates the (logged and centred) relative (*rel.freq*) and absolute (*abs.freq*) frequency of the word + accent pair, and the *coda type* as fixed effects. The correlation of the fixed effects was very low ($r < 0.1$ as computed by R's *lme4* package, cf. Bates et al., 2013). The model has a random intercept for *word*.

$$dist \sim rel.freq + abs.freq + coda.type + (1|word) \tag{4}$$

An overview of the model's fixed effects is given in Table 4. As outlined above, the intercept corresponds to the baseline condition where all factors have the "default" value, to which the effects of the other factors are added. The baseline for the model was the value predicted for $rel.freq = 0$, $abs.freq = 0$, $coda.type = -V$ (unvoiced). The coefficient of each significant predictor gives the deviation from the baseline. For the numerical frequency values it indicates the change with each unit increase in frequency, for the categorical value *coda type*, it indicates the deviation for the given level of the variable (i.e. for sonorant (+*S*), or for voiced, but not sonorant (+*V–S*)).

Table 4
Estimated coefficients, SE, and *t*-Value for the linear mixed model predicting average distance for H*L dataset in experiment 2.

	Estimate	SE	*t*-Value
(Intercept)	3.92	0.15	26.81
rel.freq	−0.30	0.09	−3.30
abs.freq	0.36	0.10	3.52
coda.type + S	−0.60	0.16	−3.66
coda.type + V–S	−1.91	1.58	−1.21

The parameter of interest, the relative frequency with which a word type occurs with an H*L accent, affects the average distance, and thus the variability of the pitch accent tokens, significantly ($\chi^2 = 10.766, p < 0.005$ compared to the respective null model). The coefficient is negative (cf. Table 4), that is, with increasing relative frequency of a word + accent pair, the variability of the accent tokens decreases (reflected by a decrease in average Euclidean distance).

Interestingly, absolute frequency was also a significant predictor of pitch accent variability with the opposite effect: while relative frequency lowers pitch accent variability, absolute frequency had an increasing effect ($\chi^2 = 12.23, p < 0.0005$ compared to the respective null model).

Coda type was a significant predictor as well ($\chi^2 = 14.063, p < 0.001$) with sonorant codas decreasing the variability. The effect for codas that are voiced, but not sonorant is not significant ($t = -1.21$, corresponding to $p \approx 0.22$, cf. Table 4).

*2.3.3.2. L*H accents.* As for the L*H dataset, the model which fitted the data the best was characterised by a model term (given in Eq. (5)) including only the absolute frequency of word + accent pair as a fixed factor and a random intercept for *word*.

$$dist \sim abs.freq + (1|word) \tag{5}$$

That is, contrary to our hypothesis, the model did not include the relative frequency of a word + accent pair. Absolute frequency, however, showed the same effect as observed in the H*L dataset: with increasing absolute frequency of word + accent pair, pitch accent variability increases as well ($\beta = 0.25 \pm 0.07$ SE, $\chi^2 = 14.917, p < 0.0005$ compared to the null model). The discussion (Section 3) offers some possible explanations for the absence of the effect of relative frequency for L*H.

2.3.3.3. Summary. The results from experiment 2 demonstrate further that the absolute frequency with which the pitch accent and the word occur together influences pitch accent shape. Moreover, the effects of relative frequency found in the H*L dataset indicate that relative frequency is also an influential factor for pitch accent shape realisation. The greater the relative frequency, i.e. the greater the proportion of tokens where a specific word occurs with H*L (compared to all pitch accented instances of the word), the smaller the variability in the realisation of the H*L accents.

Again, these results highlight that intonation is affected by characteristics of the word level and the distributional properties of the combined types. The implications for theories of pitch accent assignment are discussed in Section 3.

2.4. Experiment 3: Relative frequency of word in lexical context

Experiments 1 and 2 demonstrated that distributional properties of combinations of pitch accents and words influence the acoustic detail of pitch accent realisation, thus supporting the hypothesis of a relationship between the lexical and the tonal level. Our third experiment looked at the relative frequency of a word given its left and right neighbour. In Schweitzer et al. (2011) we showed that relative word frequency in trigram contexts affects the variability of the prosodic context in which a word occurs as well as the variability of pitch accents on that word, if it is accented. However, here, we use linear mixed effects models to replicate these findings, as mixed models avoid some problematic aspects that arise when using regular linear regression models on our data. This experiment adds another dimension to the ways in which the lexical level of linguistic structure influences the tonal level: in this experiment, frequency effects on the lexical level alone (as opposed to the frequency of word-pitch accent combinations in experiments 1 and 2) are shown to influence tonal realisation.

Pitch accent placement has been shown to be dependent on relative word frequency (Pan and Hirschberg, 2000) and it has been argued that *idioms* are expected to occur with a narrow range of tonal contours (Bolinger, 1985), but the detailed prosodic properties of frequent lexical sequences have, to our knowledge, not been investigated before our experimental work.

2.4.1. Data

In this experiment, a subset of the Switchboard corpus was analysed: it consists of a collection of spontaneous telephone conversations between American English speakers (Godfrey et al., 1992). 76 of the conversations, or around 6 h of speech from 114 speakers, are annotated for pitch accent location and prosodic boundary location (Calhoun et al., 2010) according to the ToBI standard (Beckman and Hirschberg, 1999). Pitch accent type is not marked.

2.4.1.1. Prosodic realisation. For the two analyses presented here, two different datasets were extracted. For the examination of prosodic pattern variability, trigrams that occurred *at least 4 times* in the prosodically annotated part of Switchboard were extracted. Trigrams involving some hesitation fillers were set aside, resulting in a dataset of 3705 tokens (124 word types of the middle word in the trigram, which occurred in 541 trigram types – both trigrams and words ranged from 4 to 50 tokens per type). We will refer to this dataset as the *prosodic pattern dataset*.

For the investigation of pitch accent variability, we extracted only trigrams that occurred with an accent on the middle word. Since the calculation of pitch accent variability involves some data processing which reduced the data (e.g. outlier removal and exclusion of pitch accents which could not be modelled with one or two sigmoids, see below), we restricted the analysed data to those trigrams which occurred *at least 4 times with an accent on the middle word* in the reduced dataset (400 tokens, 67 trigram types, 34 word types with the word types ranging from 4 to 407 tokens per type, and the trigrams ranging from 4 to 93 tokens per type; the dataset before applying these restrictions comprised 649 tokens, 104 trigram types and 48 word types). This dataset will be called the *pitch accent variability dataset* in the following.

2.4.1.2. Lexical frequency. To calculate relative word frequency, word and trigram frequencies were extracted from the whole Switchboard corpus, along with the Callhome American English corpus (Kingsbury et al., 1997), a smaller corpus of spontaneous telephone conversations. The combined corpus comprised just over three million words.

The frequency of each trigram was divided by the frequency of the middle word (in any trigram context) in the combined corpus; the resulting value is the probability of that word in that trigram context. Table 5 lists the trigrams with the highest relative word frequency in Switchboard and Callhome for the two analysed datasets. Most of these would be considered collocations in English, affirming that our relative frequency measure is working as intended.

2.4.2. Methodology

2.4.2.1. Calculation of prosodic context variability. To capture the prosodic context of a word, we determine for each trigram token its prosodic pattern. To this end, each word in a trigram was classified as being accented or not, and as carrying a boundary or not. Then, the prosodic pattern of the token was given by the sequence of classifications of the three words. For example, for the trigram *a lot of*, if there was an accent on *lot* and a boundary after *of*, the prosodic pattern of the word sequence *a—lot—of* was encoded as *NoAcc-NoBound—Acc-NoBound—NoAcc-Bound*.

For each trigram type (e.g. the type "a lot of") we determined the most common prosodic pattern in which the middle word occurred. For example, the word "lot" occurred in 10 different prosodic patterns in the prosodic pattern dataset. The most common pattern, i.e. the pattern in which "lot" occurred most often was *NoAcc-NoBound—Acc-NoBound—NoAcc-Bound*. We then determined for each trigram token with "lot" as a middle word, whether or not the trigram was realised with the prosodic pattern *NoAcc-NoBound—Acc-NoBound—NoAcc-Bound*.

If a large portion of the trigram tokens for a given word are realised with one dominant prosodic pattern, the coupling between the word and its prosodic context is strong, i.e. there is little variability in prosodic realisation. On the other hand, if for a given word the most common prosodic pattern on its trigram tokens is still relatively infrequent, this means that there is no dominant prosodic pattern for that word in its trigram context; rather, there is high variability in the prosodic realisation of that word.

2.4.2.2. Calculation of pitch accent variability. The calculation of pitch accent variability was analogous to the methodology employed in experiment 2. For all the accent tokens, PaIntE parameters were extracted. Since the data was not annotated for pitch accent types, we used all accents that could be approximated using one or two sigmoids. Outlying tokens were removed analogously to experiment 1 and 2, i.e. outliers were more than 1.5 interquartile range (IQR) away from the quartiles. In one case the PaIntE function returned a negative value for *c*2, this pitch accent token was removed. Then we determined the pitch accent variability as described in Section 2.3.2.1. However, this time, for each token of a *trigram* type with an accent on the middle word, the average distance to all other tokens of the same trigram type was determined.

2.4.2.3. Statistical analysis. The prosodic pattern dataset and the pitch accent variability dataset were again analysed using the *lme4* package in R (Bates et al., 2013).

For the prosodic pattern analysis, we fitted a generalised linear mixed model using the logit link function. The dependent variable was the binary *most.common.intonation*, indicating whether or not the trigram was realised with the most common prosodic pattern in which the middle word occurs. We tested the logged and centred relative lexical frequency in the large Callhome/Switchboard corpus, the logged and centred absolute word frequency (to control for effects of absolute frequency of the unigram) and the logged and centred trigram frequency as fixed effects. As random effects, we tested the speaker, the word, and the trigram.

For the pitch accent variability analysis, we fitted a linear mixed model using variability as the dependent variable. The logged and centred relative lexical frequency was tested as a fixed effect, as was the position of the word in the phrase and the logged and centred frequencies of word and trigram. As random effects, we tested the trigram and the (middle) word. Note that, as in experiment 2, we did not have to include speaker as a random factor, since our methodology for measuring the pitch accent variability uses *z*-scored PaIntE values and therefore normalises for speaker differences (cf. Section 2.3.2.1).

Table 5
Trigram types with the highest relative word frequency of the middle word in the combined Switchboard and Callhome in the two analysed datasets.

	Prosodic pattern dataset		Pitch accent variability dataset	
	P_{lex}	Trigram	P_{lex}	Trigram
1	0.738	the rest of	0.738	the rest of
2	0.659	a lot of	0.659	a lot of
3	0.609	I grew up	0.544	as far as
4	0.544	as far as	0.484	to worry about
5	0.541	be able to	0.432	a matter of
6	0.494	a couple of	0.279	I don't know
7	0.484	to worry about	0.278	I used to
8	0.432	a matter of	0.272	it seems like
9	0.4	a nursing home	0.268	a little bit
10	0.397	as soon as	0.164	take care of
11	0.279	I don't know	0.108	I guess I
12	0.278	I used to	0.106	the point where
13	0.272	it seems like	0.094	it's kind of
14	0.268	a little bit	0.093	that kind of
15	0.217	I ended up	0.092	and stuff like

Analogously to experiment 1 and 2 we determined whether each factor significantly improved the model using likelihood ratio tests (Baayen, 2008; Baayen et al., 2008; Winter, 2013). The *p*-values reported for the factors below were obtained by comparing the null models to the models with the respective factor included.

2.4.3. Results

2.4.3.1. Prosodic context variability. For the prosodic pattern analysis, the model term for the generalised mixed model fitting the data best is given in Eq. (6). The model incorporates the (logged and centred) relative lexical frequency in the combined corpus (*rel.word.freq*)) and the (logged and centred) absolute frequency of the middle word (*word.freq*) as fixed effects. As random effects, the model has intercepts for *word*, *speaker* and *trigram*.

$$most.common.intonation \sim rel.word.freq + word.freq + (1|word) + (1|speaker) + (1|trigram) \quad (6)$$

An overview of the model's fixed effects is given in Table 6. As the relative lexical frequency increases, the likelihood of the most common prosodic pattern for that word in its trigram context increases significantly ($\chi^2 = 208.63, p \ll 0.0001$). An increase in absolute word frequency decreases this likelihood ($\chi^2 = 14.069, p < 0.0005$). That is, similarly to the result for H*L in experiment 2, an increase in relative frequency yields a decrease in prosodic variability, whereas an increase in absolute frequency affects variability in the opposite direction.

Note that in this model, the estimators of the two fixed effects were correlated ($r = 0.67$, as computed by R's *lme4* package, cf. Bates et al., 2013), which might indicate collinearity. Though their variance inflation factor ($VIF = 1.8$) and the condition number ($\kappa = 2.6$) are uncritical (Myers, 1990; Menard, 1995), we fitted a linear regression model which predicted word frequency on the basis of relative word frequency. We then replaced word frequency by the residuals of the regression as a predictor in the linear mixed model. In the resulting model, correlation of the fixed effects was very low ($r = -0.09$ as computed by *lme4*, cf. Bates et al., 2013). An overview of this corrected model is given in Table 7. Comparing it to the respective null models results in significance for the effect of relative word frequency ($\chi^2 = 38.179, p \ll 0.0001$) and for the effect of word frequency ($\chi^2 = 14.069, p < 0.0005$).

As the relative lexical frequency increases, the likelihood of the most common prosodic pattern for that word in its trigram context increases significantly. The intercept of −1.16 is the value the model would predict for cases where the logged and centred relative frequency is 0. This hypothetical value corresponds to a raw relative frequency value of 0.013. The predicted value then gives the log-odds of being accented with the most common intonation pattern, and for the intercept of −1.16 this corresponds to a probability of approx. 0.24. For the lowest relative frequency in our data (0.000046) the model predicts a log-odds of −2.71, or a probability of 0.06, of being accented with the most common intonation pattern. For the highest relative frequency in our data (0.74) the model predicts a log-odds of −0.05, or a probability of 0.49, of being accented with the most common intonation pattern.

2.4.3.2. Pitch accent variability. For the pitch accent variability analysis, the model term for the model fitting the data best is given in Eq. (7). Relative lexical frequency (*rel.word.freq*) is a fixed effect, and *trigram* a random effect.

$$dist \sim rel.word.freq + (1|trigram) \quad (7)$$

An overview of the fixed effects in the model is given in Table 8. Relative lexical frequency affects pitch accent variability (i.e. *average distance*) significantly ($\chi^2 = 19.233, p \ll 0.0001$), in that with increasing relative frequency pitch accent variability decreases.

2.4.3.3. Summary. The results in experiment 3 add further evidence that realisation at the tonal level is influenced by distributional properties at the lexical level. In addition to probabilistic aspects of the combination of a specific pitch accent with a specific word (demonstrated in the first two experiments), probabilistic factors of the word level alone influence intonation. Both parameters measuring prosodic variability, on or around the word, indicate that prosodic variability decreases as the probability of a word in its (lexical) context increases.

While autosegmental theories of intonation are silent on such frequency effects, exemplar models would in fact expect them. Exemplars are assumed to contain detailed acoustic information, hence they are likely to also contain the fundamental frequency contour. The implications of these findings are set out in more detail in the following.

Table 6
Estimated coefficients, SE, and *z*-Value for the generalised linear mixed model predicting occurrence with most common prosodic context in experiment 3.

	Estimate	SE	*z*-Value
(Intercept)	−1.16	0.11	-10.13
rel.word.freq	0.11	0.06	1.96
word.freq	−0.31	0.08	−3.87

Table 7
Estimated coefficients, SE, and *z*-Value for the corrected generalised linear mixed model predicting occurrence with most common prosodic context in experiment 3.

	Estimate	SE	*z*-Value
(Intercept)	−1.16	0.11	−10.13
rel.word.freq	0.27	0.04	6.44
Residuals (word.freq ∼ rel.word.freq)	−0.31	0.08	−3.87

Table 8
Estimated coefficients, SE, and *t*-Value for the linear mixed model predicting pitch accent variability in experiment 3.

	Estimate	SE	*t*-Value
(Intercept)	4.25	0.12	34.88
rel.word.freq	−0.38	0.08	−4.72

3. Discussion and conclusion

Our experimental work aimed to test whether there is evidence for interactions between prosody and the lexicon, i.e., whether prosody and words are stored together, as would be anticipated by an exemplar-theoretic view of lexical storage. Specifically, the experiments presented here sought to determine whether there are lexical frequency of occurrence effects on tonal parameters, consistent with the storage of intonational information in the lexicon. An overview of our findings is given in Table 9.

Experiment 1 looked at types of pitch accents on specific words (such as "Berlin + H*L") and demonstrated that the tonal realisation of pitch accents is sensitive to the frequency with which the pair occurs. The results demonstrated that the accent range (amplitude of the rise for rising accents, amplitude of the fall for falling accents) subtly but significantly increases with increasing frequency of the combined types. This relationship between the word level and the tonal level has no obvious explanation in autosegmental-metrical theories of intonation. Theories of episodic storage, on the other hand, expect linguistic units to be stored as concrete, highly specified instances. Therefore, properties of the lexical level are expected to be interwoven with the tonal level.

The direction of the effect may at first seem surprising, as lexical frequency has previously been shown to be linked to phonetic reduction, as outlined in Section 1.1. However, the data we analyse are crucially different from the studies mentioned above: we look at pitch accents, which are mostly assigned to put emphasis on words and draw the listeners' attention to them. Ernestus (2014) offers a listener-driven account of effects of phonetic reduction: "Speakers would like to reduce as much as possible in their articulatory effort, but only reduce those units that can easily be recognised by the listener[...]" (p.5; see also Lindblom, 1990). From this perspective, it would not be expected that words which receive a pitch accent are phonetically reduced. Rather, we believe that basic assumptions of exemplar models can offer an explanation: during speech production, the exemplars that match the intended utterance best are employed in constructing the production target. Therefore, the exemplars matching the communicative goal best are selected. In this case, the communicative function of the intended utterance requires "prominence", so the selection is among stored tokens with increased prominence, i.e. pitch accented tokens. The new instance is produced involving production noise (due to imprecision inherent in the production process) causing it to be slightly more or less prominent than the production target, and is then stored in memory again. If the production noise caused the exemplar to be realised with increased acoustic prominence, the new exemplar is likely to be selected as forming the production target in future productions when the communicative function "prominence" is required. Therefore, with increasing frequency, exemplars would be expected to be realised with slightly increased acoustic prominence. This behaviour would then be expected to be entrenched (cf. Pierrehumbert, 2001) over time, in order to avoid excessive prominence.

Experiment 2 then looked at the variability in pitch accent shape among the tokens of pitch word + accent types and related their similarity to the relative frequency of that type, controlling for the effects of absolute frequency. For H*L accents, relative frequency had an effect on the variability of the accent tokens. The more often a word type and H*L occurred together in the analysed data, relative to the occurrence of the word with any pitch accent type, the less variable the realisations of the H*L tokens. Moreover, for both accent types, H*L and L*H, an effect of the absolute frequency of the word + accent pairing was found in that greater frequency entailed greater variability. Again, these findings demonstrate an interdependence between the lexical and the tonal level with details in the tonal realisation being influenced by distributional properties involving the lexical items.

These effects would be expected in an exemplar-theoretic model which assumes storage of tonal features with words:

Table 9
Overview of the experiments and their outcomes.

	Experiment 1	Experiment 2	Experiment 3
Unit	PA + word	PA + word	word + context
Frequency measure	Absolute	Absolute and relative	Relative
Tonal parameter	PA shape	PA shape variability	PA shape variability prosodic context variability
Language	German	German	American English
Genre	Radio	Radio	Spontaneous
Result	Greater accent amplitudes with increasing frequency	Greater variability with increasing absolute frequency, less variability with increasing relative frequency for H*L	Less variability with increasing relative frequency

While generally greater frequency in the stored exemplar cloud entails greater variability (with the individual exemplars coming from different contexts, speakers, situations, dialects etc.), the relative frequency with which an accent contour and a word occur, has the opposite effect: when pitch accented instances of a word are selected to form a production target, the intended production has a particular communicative goal which entails a specific tonal contour (marked as a specific pitch accent type in the data). If the word was often realised with a communicative function entailing the same accent (hence, the relative frequency of accent type and word is high) then the set of exemplars forming the production target is more homogeneous than in the case of low relative frequency. Therefore, the new token is more likely to be produced similarly to the existing tokens of the specific accent than in the low relative frequency case. Consequently for high relative frequency cases the exemplars are expected to be more similar to each other than for low frequency cases.

In this study, however, we only found an effect for H*L accents, not for L*H. One possible explanation for this is difficulties with the PaIntE parametrisation of the pitch accent tokens. Recall that PaIntE attempts to fit two sigmoids to each accent, one representing the accent rise and the other the fall; however, if a model with either a rising or a falling sigmoid fits the accent shape better, then PaIntE reverts to a one-sigmoid based parametrisation. Our methodology aimed to make the one-sigmoid cases comparable to the two-sigmoid cases, however it remains the case that there is less variation among the one-sigmoid cases since essentially two dimensions (the steepness and range of the unused sigmoid) do not contribute to the calculation of variability. In the analysed L*H dataset, a much greater percentage of the tokens was approximated with only one sigmoid (54% one-sigmoid cases, as opposed to 19% for H*L). It is possible, therefore, that the greater proportion of one-sigmoid cases in the L*H set might have reduced the variability in the dataset sufficiently to mask any frequency effect. Investigating the one-sigmoid cases more closely, will be the subject of future work.

Experiment 3 completed the picture by investigating how relative *word* frequency influences pitch accent shape. It was found that the more probable a word in its lexical context, the more homogeneous the prosodic contexts it occurs in. Further, the shape of pitch accents on probable words (in a trigram context) turned out to be less variable than on words with a low probability. Together, the two findings demonstrate that an increase of lexical probability in a trigram context correlates with a decrease of tonal variability. This presents further evidence for cohesion between the word and its prosodic realisation. Within an exemplar framework, it is expected that words which collocate together will be stored together, and may acquire particular phonetic characteristics that those words do not have in other contexts (Hay and Bresnan, 2006). We have shown that those characteristics include prosodic properties.

The relationship between the lexical level and the tonal level revealed by our analyses makes it unlikely that prosodic realisation is solely determined by a combination of 'top-down' syntactic, semantic and pragmatic factors (e.g. given/new status), and the phonological context (e.g. syllable structure, or how close together accents are). While these factors are undoubtedly relevant, the robust relationship found between low prosodic contour variability and high relative lexical frequency (Experiment 3) strongly suggests that sequences of tonal events (pitch accents and boundary tones) are stored with lexical sequences (at least trigrams). The choice of words then seems to directly influence the choice of prosodic contour as well, rather than this being determined solely by syntactic, semantic and pragmatic factors; though more research is needed to determine exactly how lexical and syntactic/semantic/pragmatic factors interact. Furthermore, our results show that the *acoustic detail* of a given tonal contour is influenced by the distributional properties of the word level, albeit subtly: absolute frequency influences the accent range of pitch accents (Experiment 1), and relative frequency influences pitch accent shape variability (Experiments 2 and 3). These effects on pitch accent realisation are reminiscent of the effects of "word-specific phonetics" reported by Pierrehumbert (2002). That is, a given phonological entity (in this case a tonal event) varies in systematic ways depending on the distributional properties of the word it occurs on. Hence, one might even speak of "word-specific prosody" which adds to other, known effects of phonological context influencing prosodic realisation.

The question then arises of exactly what prosodic information is stored with words, and how this affects what is selected in production. One option is that word tokens and accent type tokens are stored separately, but are co-indexed if they occurred together. The other option is that the particular pitch contour is stored with the word (or collocation) as a single unit, conceivably also with the pragmatic function, i.e. the contextualised meaning, of that unit. In case of the former option, the observed effects on the acoustic detail of the realisation of accents could theoretically arise from the co-indexing of accent types with particular words. Co-indexing would imply that there is a link between an accent type token and a word token, and selecting either one would increase the likeliness of selecting the other. Take a word, e.g. *yeah*, produced with an L*+H accent, and used as an uncertain affirmation. It could be that a speaker selects the word, and then the accent type (L*+H) on the basis of its pragmatic function. Then because originally co-occurring instances of *yeah* and L*+H are indexed together, the speaker is more likely to select an L*+H token that originally occurred on *yeah* than an L*H that occurred elsewhere. In this way, storage could impact upon the acoustic detail of the accent on that word.

However, we think it is more likely, and consistent with our results, that the word(s), contour and pragmatic function are stored as a combined unit if they co-occur frequently enough. In our scenario, the speaker would wish to convey uncertain affirmation and on that basis select a combined unit of *yeah* + *L*+H-like* contour. The pitch contour on this word would display entrenched intonation because of the frequency of the combination (cf. Calhoun and Schweitzer, 2012). Thus, our results suggest that pragmatic function is part of the selection criterion of lexical exemplars in the production process (cf. Pierrehumbert, 2001), because tonal parameters are part of the exemplar representation.

It should be acknowledged, however, that the effects presented here seem to indicate a relatively subtle effect of lexical frequency on intonational realisation; though the effects are statistically significant. We think that this is because storage of intonational information is very uneven across the lexicon, so that the size of the effect might not be easy to show across a range of words as wide as tested in our models. This is supported by results reported by Calhoun and Schweitzer (2012), where effects of lexical storage of intonational information were shown to be strong for certain types of words, e.g. adverbs and discourse markers, but less clear for other types of words, e.g. concrete nouns.

It is possible, though we believe unlikely, that the results found here arise by coincidence, not co-storage and co-selection. That is, to take the example above, if the word *yeah* is often used with a particular pragmatic function, then it would be likely to be produced with a particular intonation type. This means that the tokens of *yeah* in the database would be likely to show less prosodic variability than a word with a less stable pragmatic function, without necessarily implying cognitive storage of the word with the contour. We believe that this approach offers a much less satisfactory explanation of our findings, however. It is now well established within the literature on Exemplar Theory that phonetic detail can be stored with words, and pragmatic function with phrases, and that these have effects on usage that cannot be explained without assuming storage (Bybee, 2006); there is no obvious reason why intonation should not work the same way. Among the results presented here, it is particularly hard to see why the first set of results should follow by coincidence: there is no obvious reason why frequent pairings of words with a particular accent type should have a pragmatic function that leads to them being realised with increased pitch amplitude. On the other hand, there is a natural explanation in the exemplar approach. We are looking, however, into possible methodologies that could test this directly. For example, we could extend the iterative mimicry study by Braun et al. (2006) to see whether the type of intonation attractors changes depending on the frequency of the words in the phrase.

Our results demonstrate effects on intonation that should be considered in exemplar-theoretic models. In addition to the findings indicating that tonal features should be considered part of the exemplar representation, the evidence presented here also indicates that pragmatic functions are selection criteria and that entrenchment should be modelled for tonal parameters. However, although our results provide evidence for the lexicalised storage of intonation, we are not claiming that pitch contour is solely accessed through the lexicon. Nevertheless, for traditional approaches to intonation, our data indicate that it is crucial that frequency of occurrence effects are acknowledged in order to attain a comprehensive picture of the production of intonation.

Acknowledgements

The research presented here was funded by the German National Science Foundation (DFG) within the collaborative research centre 732 (K. Schweitzer, M. Walsh, and A. Schweitzer). Part of the research conducted here was carried out while S. Calhoun was a British Academy Postdoctoral Fellow at the University of Edinburgh.

We would like to thank Janet Pierrehumbert for useful discussions on the interpretation of the results and two anonymous reviewers for their helpful comments. All inaccuracies and misinterpretations in the paper are of course our own.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki F. (Eds.), 2nd International Symposium on Information Theory, pp. 267–281.

Baayen, H., 2008. Analyzing Linguistic Data. Cambridge University Press.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412.

Bates, D., Maechler, M., Bolker, B., 2013. lme4: Linear mixed-effects models using S4 classes. <http://CRAN.R-project.org/package=lme4>, r package version 0.999999-2.

Baumann, S., Grice, M., Benzmüller, R., 2001. GToBI – a phonological system for the transcription of German intonation. In: Puppel, S., Demenko, G. (Eds.), Prosody 2000. Speech Recognition and Synthesis. Adam Mickiewicz University, Faculty of Modern Languages and Literature, Poznan, pp. 21–28.

Beckman, M., Hirschberg, J., 1999. The ToBI Annotation Conventions. <http://www.ling.ohio-state.edu/tobi/ame_tobi/annotation_conventions.html>.

Black, A.W., 1997. The Festival speech synthesis system. <www.cstr.ed.ac.uk/projects/festival.html>.

D.L. Bolinger, Intonation and its parts: Melody in spoken English, Arnold, London, 1985.

Braun, B., Johnson, E.K., 2011. Question or tone 2? How language experience and linguistic function guide pitch processing. J. Phonet., 585–594. http://dx.doi.org/10.1016/j.wocn.2011.06.002.

Braun, B., Kochanski, G., Grabe, E., Rosner, B.S., 2006. Evidence for attractors in English intonation. J. Acoust. Soc. Am. 119 (6), 4006–4015, <http://link.aip.org/link/?JAS/119/4006/1>.

Braun, B., Dainora, A., Ernestus, M., 2011. An unfamiliar intonation contour slows down online speech comprehension. Lang. Cogn.

Process. 26 (3), 350–375. http://dx.doi.org/10.1080/01690965.2010.492641, <http://www.ingentaconnect.com/content/psych/plcp/2011/00000026/00000003/art00002>.

Brenier, J.M., Nenkova, A., Kothari, A., Whitton, L., Beaver, D., Jurafsky, D., 2006. The (non)utility of linguistic features for predicting prominence in spontaneous speech. In: IEEE/ACL 2006 Workshop on Spoken Language Technology, pp. 54–57.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer.

Bybee, J., 2000. The phonology of the lexicon: evidence from lexical diffusion. In: Barlow, M., Kemmer, S. (Eds.), Usage-Based Models of Language. CSLI, Stanford, pp. 65–85.

Bybee, J., 2006. From usage to grammar: the mind's response to repetition. Language 84, 529–551.

Bybee, J., Scheibman, J., 1999. The effect of usage on degrees of constituency: the reduction of do not in English. Linguistics 37 (4), 575–596.

Calhoun, S., Schweitzer, A., 2012. Can intonation contours be lexicalised? Implications for discourse meanings. In: Elordieta Alcibar, G., Prieto, P. (Eds.), Prosody and Meaning (Interface Explorations). Mouton DeGruyter, pp. 271–328.

Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., Beaver, D., 2010. The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. Lang. Resour. Eval. 44 (4), 387–419.

Carreiras, M.A., Perea, M.B., 2004. Naming pseudowords in Spanish: effects of syllable frequency. Brain Lang. 90, 393–400.

Cholin, J., Levelt, W.J.M., Schiller, N.O., 2006. Effects of syllable frequency in speech production. Cognition 99 (2), 205–235. http://dx.doi.org/10.1016/j.cognition.2005.01.009.

Daelemans, W., Gillis, S., Durieux, G., 1994. The acquisition of stress: a data-oriented approach. Comput. Linguist. 20 (3), 421–451, <http://dl.acm.org/citation.cfm?id=204915.204923>.

Eckart, K., Riester, A., Schweitzer, K., 2012. A discourse information radio news database for linguistic analysis. In: Chiarcos, C., Nordhoff, S., Hellmann, S. (Eds.), Linked Data in Linguistics, . In: Representing and Connecting Language Data and Language Metadata. Springer, Heidelberg, pp. 65–75, ISBN 978-3-642-28248-5.

Ernestus, M., 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. Lingua 142, 27–41.

Godfrey, J., Holliman, E., McDaniel, J., 1992. Switchboard: telephone speech corpus for research and development. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92, vol. 1, pp. 517–520.

Goldinger, S.D., 1997. Words and voices—perception and production in an episodic lexicon. In: Johnson, K., Mullennix, J.W. (Eds.), Talker Variability in Speech Processing. Academic Press, San Diego, pp. 33–66.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11 (1), 10–18. http://dx.doi.org/10.1145/1656274.1656278 (ISSN 1931-0145).

Hay, J., Bresnan, J., 2006. Spoken syntax: the phonetics of giving a hand in New Zealand English. Linguist. Rev. 23, 321–349.

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2010. IMS German Festival Home Page. <www.ims.uni-stuttgart.de/phonetik/synthesis>.

Jilka, M., Möbius, B., 2007. The influence of vowel quality features on peak alignment. In: Proceedings of Interspeech 2007 (Antwerpen), 2621–2624.

Johnson, K., 1997. Speech perception without speaker normalization: an exemplar model. In: Johnson, K., Mullennix, J.W. (Eds.), Talker Variability in Speech Processing. Academic Press, San Diego, pp. 145–165.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W.D., 2001. Probabilistic relations between words: evidence from reduction in lexical production. In: Bybee, J., Hopper, P. (Eds.), Frequency and the Emergence of Linguistic Structure. Benjamins, Amsterdam, pp. 229–254.

Kingsbury, P., Strassel, S., McLemore, C., McIntyre, R., 1997. CALLHOME American English Transcripts, Linguistics Data Consortium, Philadelphia, No. LDC97T14.

Kruschke, J.K., 1992. ALCOVE: an exemplar-based connectionist model of category learning. Psychol. Rev. 99 (1), 22–44.

Ladd, D., 2008. Intonational Phonology, second ed. Cambridge University Press, Cambridge, UK.

Lee, S., Potamianos, A., Narayanan, S., 1999. Acoustics of children's speech: developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. 105 (3), 1455–1468.

Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modelling, . In: NATO ASI Series, vol. 55. Springer Netherlands, pp. 403–439. http://dx.doi.org/10.1007/978-94-009-2037-8_16, ISBN 978-94-010-7414-8.

Losiewicz, B.L., 1992. The Effect of Frequency on Linguistic Morphology. Ph.D. Thesis, University of Texas, Austin, TX.

Luce, P.A., 1986. Neighborhoods of Words in the Mental Lexicon. Ph.D. Thesis, Indiana University, Bloomington, Dept. of Psychology.

Mandel, D.R., Jusczyk, P.W., Nelson, D.G.K., 1994. Does sentential prosody help infants organize and remember speech information? Cognition 53 (2), 155–180, <http://www.sciencedirect.com/science/article/pii/0010027794900698>.

Marsi, E., Reynaert, M., van den Bosch, A., Daelemans, W., Hoste, V., 2003. Learning to predict pitch accents and prosodic boundaries in Dutch. In: Proceedings of the ACL-2003 Conference, Sapporo, Japan, pp. 489–496.

Mayer, J., 1995. Transcribing German Intonation – The Stuttgart System. Tech. Rep., Universität Stuttgart. <http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html>.

Medin, D.L., Schaffer, M.M., 1978. Context theory of classification learning. Psychol. Rev. 85 (3), 207–238, <http://www.sciencedirect.com/science/article/B6X04-4NN6WB6-5/2/cb2dda0c11adb6a6beef3ce1b7dc2058>.

Menard, S., 1995. Applied Logistic Regression Analysis. Sage, Thousand Oaks, CA.

Möhler, G., 2001. Improvements of the PaIntE Model for F_0 Parametrization. Tech. Rep., Institute of Natural Language Processing, University of Stuttgart, draft version.

Möhler, G., Conkie, A., 1998. Parametric modeling of intonation using vector quantization. In: Proceedings of the Third International Workshop on Speech Synthesis (Jenolan Caves, Australia), pp. 311–316.

Mücke, D., Grice, M., Becker, J., Hermes, A., Baumann, S., 2006. Articulatory and acoustic correlates of prenuclear and nuclear accents. In: Proceedings of Speech Prosody 2006 (Dresden), pp. 297–300.

Myers, R.H., 1990. Classical and Modern Regression with Applications, seconf ed., Duxbury, Boston, MA.

Nenkova, A., Brenier, J., Kothari, A., Calhoun, S., Whitton, L., Beaver, D., Jurafsky, D., 2007. To memorize or to predict: prominence labeling in conversational speech. In: Proceedings of NAACL-HLT, pp. 9–16.

Nosofsky, R.M., 1986. Attention, similarity, and the identification–categorization relationship. J. Exp. Psychol.: Gen. 115 (1), 39–57.

Nosofsky, R.M., Kruschke, J.K., McKinley, S.C., 1992. Combining exemplar-based category representations and connectionist learning rules. J. Exp. Psychol.: Learn. Mem. Cogn. 18 (2), 211–233.

Pan, S., Hirschberg, J., 2000. Modeling local context for pitch accent prediction. Proceedings of the ACL-2003 Conference. Association for Computational Linguistics, Morristown, NJ, USA, pp. 233–240. http://dx.doi.org/10.3115/1075218.107524.

Pan, S., McKeown, K.R., 1999. Word informativeness and automatic pitch accent modeling. In: Proceedings of EMNLP/VLC 99, pp. 148–157.

Pierrehumbert, J.B., 1980. The Phonology and Phonetics of English Intonation. Ph.D. Thesis, Massachusetts Institute of Technology.

Pierrehumbert, J., 2000. Tonal elements and their alignment. In: Horne, M. (Ed.), Prosody: Theory and Experiment—Studies Presented to Gösta Bruce. Kluwer, Dordrecht, pp. 11–36.

Pierrehumbert, J., 2001. Exemplar dynamics: word frequency, lenition and contrast. In: Bybee, J., Hopper, P. (Eds.), Frequency and the Emergence of Linguistic Structure. Benjamins, Amsterdam, The Netherlands, pp. 137–157.

Pierrehumbert, J., 2002. Word-specific phonetics. In: Gussenhoven, C., Warner, N. (Eds.), Laboratory Phonology 7. Mouton de Gruyter, Berlin, Germany, pp. 101–140.

Pluymaekers, M., Ernestus, M., Baayen, R.H., 2005. Lexical frequency and acoustic reduction in spoken Dutch. J. Acoust. Soc. Am. 118 (4), 2561–2569. http://dx.doi.org/10.1121/1.2011150, <http://link.aip.org/link/?JAS/118/2561/1>.

R Core Team, 2013. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Schweitzer, A., 2010. Production and Perception of Prosodic Events – Evidence from Corpus-based Experiments, Doctoral dissertation, Universität Stuttgart.

Schweitzer, A., Möbius, B., 2004. Exemplar-based production of prosody: evidence from segment and syllable durations. In: Speech Prosody 2004 (Nara, Japan), pp. 459–462.

Schweitzer, K., Walsh, M., Möbius, B., Schütze, H., 2010a. Frequency of occurrence effects on pitch accent realisation. In: Proceedings of Interspeech 2010, Makuhari, Japan, pp. 138–141.

Schweitzer, K., Calhoun, S., Schütze, H., Schweitzer, A., Walsh, M., 2010b. Relative frequency affects pitch accent realisation: evidence for exemplar storage of prosody. In: Proceedings of the Thirteenth Australasian International Conference on Speech Science and Technology (SST) 2010, Melbourne, Australia, pp. 62–65.

Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., 2011. Prosodic variability in lexical sequences: intonation entrenches too. In: Proceedings of the International Congress of Phonetic Sciences 2011, Hong Kong, pp. 1778–1781.

Silverman, K., Backman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J., 1992. ToBI: a standard for labeling English prosody. In: Proceedings of the 1992 International Conference on Spoken Language Processing, vol. 2, Banff, Canada, pp. 867–870.

Van Lancker, D., Canter, G.J., 1981. Idiomatic versus literal interpretations of ditropically ambiguous sentences. J. Speech Hear. Res. 24 (1), 64–69, <http://jslhr.asha.org/cgi/reprint/24/1/64.pdf>.

Van Lancker, D., Canter, G.J., Terbeek, D., 1981. Disambiguation of ditropic sentences: acoustic and phonetic cues. J. Speech Hear. Res. 24 (3), 330–335, <http://jslhr.asha.org/cgi/reprint/24/3/330.pdf>.

Van Santen, J.P.H., Hirschberg, J., 1994. Segmental effects on timing and height of pitch contours. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), ISCA, Yokohama, Japan, pp. 719–722. <http://dblp.uni-trier.de/db/conf/interspeech/icslp1994.html#SantenH94>.

Van Santen, J.P.H., Möbius, B., 2000. A quantitative model of F0 generation and alignment. In: Botinis, A. (Ed.), Intonation—Analysis, Modelling and Technology. Kluwer, Dordrecht, pp. 269–288.

Vigário, M., Freitas, M.J., Frota, S., 2006. Grammar and frequency effects in the acquisition of prosodic words in European Portuguese. Lang. Speech 49 (2), 175–203. http://dx.doi.org/10.1177/00238309060490020301, <http://las.sagepub.com/content/49/2/175.full.pdf+html>.

Wade, T., Dogil, G., Schütze, H., Walsh, M., Möbius, B., 2010. Syllable frequency effects in a context-sensitive segment production model. J. Phonet. 38 (2), 227–239, <http://www.sciencedirect.com/science/article/B6WKT-4YVG80R-2/2/0e6e2c74dc9eb8449cb38ff406f37982>.

Walsh, M., Schütze, H., Möbius, B., Schweitzer, A. 2007. An exemplar-theoretic account of syllable frequency effects. In: Proceedings of the International Congress of Phonetic Sciences, Saarbrücken, Germany, pp. 481–484.

Walsh, M., Möbius, B., Wade, T., Schütze, H., 2010. Multi-level exemplar theory. Cogn. Sci. 34, 537–582.

Winter, B., 2013. Linear Models and Linear Mixed Effects Models in R with Linguistic Applications. arXiv:1308.5499, <http://arxiv.org/pdf/1308.5499.pdf>.

Wright, R., 1997. Lexical competition and reduction in speech: a preliminary report. In: Research on Spoken Language Processing Progress Report No. 21, Indiana University, Bloomington, Indiana, 1997, pp. 471–485.