Journal of Phonetics 41 (2013) 29–47

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Acoustic and respiratory evidence for utterance planning in German

Susanne Fuchs[a,*,1], Caterina Petrone[a,d,1], Jelena Krivokapić[b], Philip Hoole[c]

[a] Centre for General Linguistics (ZAS), Berlin, Germany
[b] Yale University, New Haven, USA
[c] Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, Germany
[d] Laboratoire Parole et Langage (LPL), CNRS UMR 7309, Aix-en-Provence, France

ARTICLE INFO

Article history:
Received 10 May 2011
Received in revised form
17 August 2012
Accepted 27 August 2012
Available online 28 November 2012

ABSTRACT

This study investigates prosodic planning in a reading task in German. We analyse how the utterance length and syntactic complexity of an upcoming sentence affect two acoustic parameters (pause duration and the initial fundamental frequency peak) and two respiratory parameters (inhalation depth and inhalation duration). Two experiments were carried out.

In the first experiment, data for twelve native speakers of German were recorded. They read sentences varying in length (short, long) and syntactic complexity (simple, complex). Data were analysed on the basis of the four phonetic parameters. Pause duration, inhalation depth and inhalation duration showed significant differences with respect to sentence length, but not to syntactic complexity. The initial f0 peak was not influenced by variations in length or syntactic complexity.

In the second experiment it was hypothesized that the initial f0 peak is only sensitive to length manipulations of the first constituent. Twenty speakers were recorded reading utterances varying in the length of the first (short, medium, long) and last syntactic constituent (short, long). Results for the initial f0 peak confirmed our hypothesis. It is concluded that the breathing parameters and pause duration are global parameters for planning of the upcoming sentence whereas the height of the fundamental frequency peak is a more local measure sensitive to the length of the first constituent.

1. Introduction

Language production is incremental: utterance generation proceeds in a piecemeal (by ‘‘increments’’) fashion, with planning at different levels of the linguistic and articulatory processes being interleaved (e.g., Dell, Burger, & Svec, 1997; Ferreira, 1996; Ferreira & Swets, 2002; Keating & Shattuck-Hufnagel, 2002; Levelt, 1989; Meyer, Belke, Häcker, & Mortensen, 2007; Roelofs, 1998; Wheeldon & Lahiri, 1997). However, the issues of how far ahead speakers plan the upcoming sentence, how flexible the units of planning are, and which parameters mirror utterance planning are far from settled. Concerning the units of planning, suggestions have ranged from the prosodic word to the intonational phrase (e.g., Ferreira, 1991; Krivokapić, 2007a; Levelt, 1989; Martin, Crowther, Knight, Tamborello, & Yang, 2010; Wheeldon & Lahiri, 1997). Studies assuming fixed units of planning crucially concentrate on different linguistic material (e.g., the number of phonological words, content words, stressed syllables in an utterance) and report their findings in relation to such linguistic variables (e.g., Wheeldon & Lahiri, 1997). Studies assuming more flexible units of planning additionally focus on task-related and speaker-specific behaviour with respect to planning parameters (Ferreira & Swets, 2002; Swets, Desmet, Hambrick, & Ferreira, 2007; Wagner, Jescheniak, & Schriefers, 2010). In the majority of cases the analyses are based on one of a limited number of parameters, such as reaction/response time, initiation time, pause duration or fixation points of eye movements (e.g., Ferreira, 1991; Lashley, 1951; Latash & Mikaelin, 2011; Meyer, 2004; Sternberg, Monsell, Knoll, & Wright, 1978).

In the current work on German we will concentrate on phonetic parameters which we expect to correspond to different aspects of speech planning and thus might be used to capture different scopes of planning. It should be pointed out that, given the incremental nature of speech production, we do not assume that speakers fully plan a sentence before speech onset, and that therefore the parameters we examine reflect only a rough rather than detailed sentence plan (see also Keating & Shattuck-Hufnagel, 2002). Our studies will focus on reading tasks. This task does not involve message generation, lexical retrieval or the selection of the syntactic construction; accordingly, the planning processes we examine differ in some aspects from planning in other styles such as in spontaneous speech. However, reading tasks have been used in many

* Correspondence to: ZAS/Phonetik, Schützenstrasse 18, 10117 Berlin, Germany. Tel.: +49 30 20192 569; fax: +49 30 20192 402.
E-mail address: fuchs@zas.gwz-berlin.de (S. Fuchs).

[1] The first two authors have made equal contributions to the paper, hence we put them in alphabetical order.

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2012.08.007

studies since they allow one to control the spoken speech material and to gain insights into the effect of linguistic factors (such as phonological length) on different parameters of planning. Specifically, we examine the following parameters:

Parameter I, *pause duration*, is associated with the time required to plan the upcoming sentence.
Parameter II, utterance-initial *fundamental frequency peak*, is a parameter which is often discussed in relation to planning the tonal space of the utterance.
Parameter III is *inhalation depth.* This clearly forms part of the articulatory plan for an utterance. The main question here is the extent to which the speaker incorporates an estimate of the amount of air required for the utterance into this plan.
Finally, we also analyse Parameter IV, *inhalation duration*. It is expected to correspond closely to pause duration, but represents the duration for the physiological activity of inhaling.

The following sections discuss first each parameter in more detail and then the potential links among them. We are interested in the extent to which these phonetic parameters are influenced by two factors: utterance/sentence length and syntactic complexity in German. Furthermore, while we focus on general trends in planning, we will also refer to speaker-specific behaviour in order to characterize the potential flexibility in speech planning.

1.1. Parameter I: Pause duration

A commonly-used indicator of utterance planning is pause duration. Pauses can be divided into breathing, non-breathing, and filled pauses. Breathing pauses correspond to pauses during speech production where speakers take a breath, non-breathing pauses are silent temporal intervals between utterances (Grosjean & Collins, 1979), and filled pauses are non-silent pauses with fillers like *uh* and *uhm* (Clark & Fox Tree, 2002; Corley & Hartsuiker, 2011).

Most research on pause duration that we are aware of focuses on breathing and non-breathing pauses. In the majority of these studies no explicit distinction between the two is made. One reason for this could be that Grosjean and Collins (1979) report that breathing and non-breathing pauses pattern in the same manner in terms of speech planning. Another reason could be that most studies focus on acoustic signals where breathing can only be defined when breathing noise is present in the signal and the quality of the acoustic recording is relatively good. Without physiological data, it may not always be clear which pauses are breathing pauses and which are silent pauses only.

We focus in this section on the literature of non-breathing pauses, since we are not aware of studies other than that of Grosjean and Collins (1979) that explicitly distinguish between non-breathing and breathing pauses. Starting with work in the nineteen-sixties (see overview of early work in Goldman-Eisler, 1968), evidence has accumulated showing a relationship between speech pauses and the length, syntactic complexity and prosodic complexity of the surrounding utterances, where the complexity of a phrase is typically defined as the number of syntactic/prosodic brackets in the phrase. Regarding syntactic complexity, Cooper and Paccia-Cooper (1980) found that pauses inside a major syntactic constituent are shorter than pauses between major constituents. In a similar vein, pauses within sentences have been found to be shorter than pauses between sentences (Grosjean, Grosjean, & Lane, 1979 for English; Butcher, 1981 for German; Sanderman & Collier, 1995 for Dutch). For various languages it has been attested that more complex syntactic structure leads to an increase in the likelihood of pause occurrence and to longer pauses (e.g., Cooper & Paccia-Cooper, 1980; Ferreira, 1991; Grosjean et al., 1979 for English; Strangert, 1991, 1997 for Swedish; Sanderman & Collier, 1995; Terken & Collier, 1992 for Dutch). Gayraud and Martinie (2008), however, report no significant differences in an investigation of different degrees of syntactic embedding in French.

Most of the studies have not necessarily been conducted with the specific goal of examining speech planning, but with the goal of predicting the occurrence and duration of pauses across an utterance. There is a potential problem in that many of the studies mentioned do not distinguish between the effects of the utterances before and after the pause (e.g., Gee & Grosjean, 1983; Gelfer, Harris, Collier, & Baer, 1983; Grosjean et al., 1979). Given that speech planning refers to the upcoming utterance and not to the preceding one (in the sense that planning is already completed for the preceding material at the time the pause starts), but pause duration can be influenced by both, such experiments do not allow for straightforward conclusions about speech planning (see also Ferreira, 2007 on this point). Some of the exceptions to this trend are the studies by Ferreira (1991), Strangert (1997) and Whalen and Kinsella-Shaw (1997). In her seminal study, Ferreira (1991) examines the effects of syntactic complexity of the subject and object constituents on sentence initiation time. In general, only the complexity of the subject had an effect, such that more complex subjects led to longer sentence initiation times. However, she reports that speakers tend to pause within the sentence, typically between a subject and the verb, when both the object and the subject are complex. She interprets this finding as indicating that when both the subject and the verb phrase are syntactically complex, speakers are not able to complete the planning of the verb phrase during the production of the subject. Thus, an additional pause often occurs between the subject and verb phrase, reflecting the speaker's need for more time to plan the upcoming verb phrase.

Utterance length is another factor that has been shown to affect planning parameters. A number of studies have found that an increase in utterance length (expressed in feet, phonological words, or syllables) leads to an increase in pause duration and sentence initiation times (Ferreira, 1991; Kentner, 2007; Krivokapić, 2007a, 2007b, 2010; Sternberg et al., 1978; Watson & Gibson, 2004; Wheeldon & Lahiri, 1997; Zvonik & Cummins, 2003).

Finally, prosodic structure has been found to have an effect on pause duration. Examining the data from Grosjean et al. (1979), Gee and Grosjean (1983) found that pause duration is better accounted for if, in addition to syntactic structure, prosodic structure is taken into account. They suggest that the unit for speech planning might coincide with the phonological phrase and find that pause duration increases with the complexity of prosodic structure. Relatedly, a study by Ferreira (1993) examined pause duration and found it to be determined by prosodic rather than by syntactic structure. Our previous work (Krivokapić, 2007a, 2007b) also examined the effect of length and prosodic complexity on pause duration. The effect was such that for shorter prosodic phrases, prosodic complexity was accompanied by increased pause duration (Krivokapić, 2007b) whereas for long prosodic phrases, prosodic complexity led to shorter pauses (Krivokapić, 2007a).

In sum, pause duration has been established as a parameter that indicates the time the speaker needs to plan an upcoming utterance. Pause duration is influenced by a number of factors, such as the length of an utterance and its syntactic and prosodic complexity. Generally, an increase in utterance length and/or in syntactic complexity leads to an increase in pause duration.

1.2. Parameter II: Initial fundamental frequency (f0) peak

The second acoustic parameter of utterance planning is the height of the utterance-initial f0-peak. The intonation literature is still divided on the role of this parameter. The idea of *f0 planning*[2] is often linked to that of declination, i.e., the gradual downward trend of fundamental frequency over the course of an utterance (Cohen & 'tHart, 1967).

A dichotomy between hard vs. soft planning was proposed by Liberman and Pierrehumbert (1984). Hard planning is seen as ''an essential part of intending to say something and that normally needs to be accomplished before executing that intention'' (Liberman & Pierrehumbert, 1984:220). In terms of f0 implementation, hard planning would predict that *all* speakers know in advance how long their utterance is going to be and they will thus systematically plan f0 declination in advance. Soft planning, on the other hand, implies that speakers plan the f0 declination only under certain circumstances.

Models integrating hard planning usually imply a large scope of planning. For instance, Cooper and Sorensen (1981) state that in English f0 declination is a global component which is systematically planned at a sentence level. Within the framework of the Autosegmental-Metrical model, however, Pierrehumbert (1980), Pierrehumbert and Beckman (1988) and Liberman and Pierrehumbert (1984) state that much of the f0 declination can be attributed to a distinct, phonologically controlled phenomenon called ''downstep'' or ''catathesis'', which applies locally to individual pitch accents (i.e., the f0 peaks or valleys which are associated to stressed syllables and signal their prominence; they are often instantiated as local f0 maxima and minima in the f0 contours). Downstep would lead the f0 peaks in a sentence to be progressively lowered relative to the previous ones. Liberman and Pierrehumbert (1984) find very little evidence for f0 planning at sentence level, thus supporting the soft planning view.

A method that is commonly used to test whether f0 contours are planned at a sentence level or at a more local level is to investigate the relationship between height of the sentence initial f0 peak and the length of the sentence. If speakers anticipate how long the sentence will be, they will start with higher initial f0 peaks when the sentence is longer and with lower initial f0 peaks when the sentence is shorter. Discrepant results, both across and within languages are found in the literature. The following studies support the idea that f0 contours are globally planned: for English (Cooper & Sorensen, 1981), Dutch ('tHart, 1979) and Danish (Thorsen, 1986). The initial f0 peak raising has also been observed in tonal languages such as Yoruba (Laniran & Clements, 2003) and Thai (Gandour, Potisuk, & Dechongkit, 1994). In some tone languages, it has been found that initial f0 raising is a systematic phenomenon, which can even be phonologized as an extra-high tone (Rialland, 2001).

Another group of studies finds that changes in sentence length had little or no impact on the sentence-initial f0 peak, for instance, in English (Liberman & Pierrehumbert, 1984), Dutch (Van den Berg, Gussenhoven, & Rietveld, 1992), Italian and Portuguese (Avesani, 1987; Prieto, D'Imperio, Elordieta, Frota, & Vigário, 2006), Mexican Spanish (Prieto, Shih, & Nibert, 1996), Catalan (Vanrell Bosch, 2007), and Mambila (Connell, 2003, 2004).

Some studies point out that the initial f0 peak is more sensitive to the length and prosodic complexity of the first constituent of the sentence. Two are of particular relevance for our work. The first one is from Ladd and Johnson (1987). They test whether length and prosodic complexity affect initial f0 raising in English. They report results from two native speakers, who produced sentences of the form subject–verb–object or subject–verb–prepositional phrase. The subject and the object (or the prepositional phrase) were modified with respect to the number of prosodic words, and thus to their length. Moreover, they tested whether possible effects on f0 raising are due to prosody or syntax by comparing these sentences with sentences manipulated as to the nature of the syntactic categories involved (such as adverb–subject–verbal phrase). It was found that speakers increase the height of the initial f0 peak as the length and prosodic complexity of the first major constituent increases, independent of its syntactic nature (e.g., subject vs. adverb). The effect of length and prosodic complexity of the sentence-final constituent was small and only found in one out of two speakers. This leads the authors to argue that initial pitch raising is sensitive to changes in the first constituents, and not to overall sentence length. The lack of syntactic effects in Ladd and Johnson (1987)'s experiment is also in line with previous data from Cooper and Sorensen (1981), who find that in sentences modified for their length (short vs. long) and syntactic complexity (presence vs. absence of a parenthetical clause), initial f0 peak values of sentences containing the parenthetical clause are very similar to first peak values of sentences without parentheticals.

Prieto et al. (2006) compare the height of utterance initial f0 peaks in relation to the length of the subject constituent in five Romance language and varieties, with the subject constituent being realized as a single prosodic constituent in subject–verb–object sentences. Their results are mixed, since some speakers show an effect of utterance length on initial f0 peak raising while others do not, and this discrepancy is found even within the same language. In line with Liberman and Pierrehumbert (1984), they argue that initial f0 planning is an optional mechanism adopted by the speaker. Generalization from these studies is limited by the small number of participants in the experiments.

The literature on f0 reviewed so far is on languages other than German, the language under investigation here. As far as we know, there are no studies testing the link between utterance initial f0 peak and prosodic/syntactic complexity and length in this language. However, intonation studies on German indicate that the height of the f0 peaks is influenced by prosodic and syntactic complexity, e.g., by the hierarchical organization among large prosodic domains and by syntactic embedding (Féry & Kentner, 2010; Féry & Truckenbrodt, 2005). Moreover, f0 planning has been found as a side-effect by Féry and Kügler (2008), who noticed that German speakers expand their register (i.e., the tonal space between f0 peaks and valleys) when producing utterances with many pitch accents. They interpret this as a strategy to accommodate downstep through long sequences of subsequent tones. Taken together, these findings on German encourage us to investigate whether initial f0 raising can be employed as a parameter of f0 planning and whether it is affected by linguistic factors such as complexity and length.

To sum up, the increase of the initial fundamental frequency peak has often been discussed as an additional planning parameter. Assuming a global planning of the f0 contour, the initial f0 peak should be higher in longer sentences than in shorter ones. The mixed results reported in the literature could be an indicator that initial f0 peak raising is an optional mechanism. Another reason for the conflicting results might be that planning the initial f0 peak is only related to the first prosodic constituent (typically it coincides with the subject) and does not span the whole sentence. Finally, speaker-specific behaviour might play a much larger role than typically assumed.

[2] Intonational studies use the terms ''f0 pre-planning'' to indicate the process of planning an f0 contour *before* producing it (Liberman & Pierrehumbert, 1984). For the sake of terminological consistency, we will refer here to ''f0 planning'' (in the same way as for ''sentence planning'') to indicate any process involving the f0 plan.

1.3. Parameters III and IV: Inhalation depth and inhalation duration

The literature on the relationship between speech planning and respiration is relatively sparse in the sense that various physiological studies have investigated parameters which could be linked to speech planning, but planning is not explicitly mentioned. Therefore the two respiratory parameters have been combined under this section.

The principal issue linking planning and respiration is the extent to which the speaker takes into account the air volume requirements of the upcoming sentence when breathing in at the start of a breath group. But speakers may also have a choice between inhaling more for a long utterance, or inhaling to the same extent, and showing differences at the end of the sentence, where they can rely on the respiratory reserve volume, i.e., the volume of air which is still in the lungs (Russell & Stathopoulos, 1998).

Conrad, Thalacker, and Schönle (1983) investigate in read speech where (at which place) speakers inhale across short sentences and how frequently inhalations occur. Speakers do not take a breath only when gas exchange requires it. The authors found that inhalation always occurs between paragraphs, is very likely at the end of a sentence, but becomes less frequent within sentences which are separated by commas or the connector ''and''.

Various authors (Grosjean & Collins, 1979; Hoole & Ziegler, 1997; Huber & Spruill, 2008; Winkworth, Ellis, & Adams, 1994) try to answer how inhalation changes with modifications in speech rate and loudness, how breathing and subglottal pressure affect the realization of single segments especially at the beginning and end of the sentence (Hoit, Solomon, & Hixon, 1993; Slifka, 2003, 2006), how respiratory patterns are shaped by reading, spontaneous speech, monologue or dialogue (Henderson, Goldman-Eisler, & Skarbek, 1965; McFarland, 2001; Winkworth, Davis, Adams, & Ellis, 1995). Only a few studies examine how differences in linguistic structure like sentence length, linguistic complexity and/or cognitive demands (Grosjean & Collins, 1979; McFarland, 2001; McFarland & Smith, 1992; Mitchell, Hoit, & Watson, 1996) affect breathing pauses.

Henderson et al. (1965) found that breathing pauses occur exclusively at syntactic junctures in reading. In contrast, in spontaneous speech one third of all breathing pauses occurred elsewhere. Grosjean and Collins (1979) were able to show for reading that the longer breathing pauses occur at major syntactic junctures whereas the shorter non-breathing pauses occur at minor syntactic breaks. Moreover, the authors varied the speech rate as another condition. With an increase in rate, breathing pauses predominated while non-breathing pauses hardly occurred any more. Grosjean and Collins (1979) concluded that pauses in general are primarily controlled by speaking rate, syntactic boundaries and to a lesser extent by sentence length.

Mitchell et al. (1996) asked their participants to speak for three minutes and varied cognitive linguistic demands. In one condition, the experimenter suggested a topic for the participants' speech (high cognitive linguistic demand) and in the second condition, the participants were able to write an outline for the same topic, i.e., they prepared their speech before acting (low cognitive linguistic demand). Differences between the two conditions were not found with respect to the general speech breathing behaviour, i.e., to lung volume, thoracic and abdominal volume. However, fluency related measures were highly sensitive to the tasks: average speaking rate was slower and number of syllables per breath group was smaller for higher cognitive demands. The results could also be interpreted with respect to the scope of planning. Under high cognitive demands, participants might be able to plan only smaller chunks per breath group whereas larger chunks might be planned for prepared speech (While this has not been systematically examined, for a discussion of this question see also Krivokapić, 2012). Under this interpretation, unprepared speech goes hand in hand with higher breathing frequency.

Hoole and Ziegler (1997) study inhalation with respect to sentence length and loudness. They report that the speaker apparently plans for the additional volume requirements of loud speech to a much greater extent than for sentence length. The information about how loud a sentence will be produced is probably available very early in the planning process since it may be part of the whole communicative situation. Hird and Kirsner (2002) write that the neural processes of planning the respiratory system ''anticipate the demands associated with future utterances….'' (p. 538). However, they focus on acoustically measured breathing pauses only.

In their study on breathing patterns in spontaneous speech, monologues and dialogues, Winkworth et al. (1995) report for all six participants that the inhalation depth correlates with the length of the following sentence. An anticipation of lung volume was found over up to seven clauses.

Like in most previous acoustic studies for pause duration, the influence of the preceding sentence was not controlled for in studies on breathing although it might have an influence. For instance, a speaker who has just realized a long preceding sentence may take a deep breath although the upcoming sentence is not long. One exception is the study by Whalen and Kinsella-Shaw (1997). Their work is based on the assumption that the production of a longer sentence needs on average more air in comparison to a shorter one. Whalen and Kinsella-Shaw instructed their participants to produce sentences on a single breath. Sentences range from 5 to 82 syllables and from 1 to 7 clauses in order to push the respiratory system to the limits of its capacity. Three speakers' breathing manoeuvres were recorded using respiratory inductance plethysmography (see the section ''Methods''). Results of the study provide evidence for a positive correlation between sentence length and inhalation duration, i.e., longer sentences coincided with a longer inhalation duration. The linear regression model was however only able to predict 8–28% of the variance in the data. Correlations between inhalation depth and utterance length showed a speaker specific pattern: a significant positive correlation with length for two speakers, but not for the third one. This study shows that speakers have the ability to plan ahead for a large chunk of speech. However, given that participants were instructed to produce each sentence in one breath, even for the extremely long sentences of 82 syllables, it seems not surprising that deeper inhalation coincided with longer sentences. It is unclear whether such effects would also be found without these instructions. In long sentences it seems also plausible that speakers would take another breath. Whalen and Kinsella-Shaw further examine the effect of syntactic complexity (measured as the number of clauses) on inhalation duration and depth and do not find any effect. However, they note that the measure of syntactic complexity used in their study is rather simple, which might explain the lack of a complexity effect. These results differ from Grosjean and Collins (1979) who examined a presumably related measure of breathing pause duration, and found that syntactic structure has an effect, such that more complex syntactic phrases are connected with longer breathing pauses.

To summarize, there are rather few studies on breathing in the context of speech planning. Inhalation depth and inhalation duration showed an effect of sentence length with deeper and longer inhalation for longer sentences. The effects of syntactic structure were less clear: while syntactic structure has an effect on the occurrence and duration of breathing pauses (Grosjean & Collins, 1979), inhalation duration and depth are not affected by it (Whalen & Kinsella-Shaw, 1997). If we assume that inhalation depth and duration are determined by similar linguistic factors as the duration of breathing pauses, one could suppose an effect of syntactic complexity on inhalation duration and depth.

1.4. Linking the parameters

Most studies on utterance planning have been based on the investigation of one or two parameters. The current work presents a first attempt to empirically test and relate four parameters, namely pause duration, initial fundamental frequency peak, inhalation depth and inhalation duration. Linking these parameters allows us to examine the relation between acoustic and physiological measures of speech planning, and to investigate how robust or sensitive each parameter is with respect to sentence length and syntactic complexity.

Pause duration and inhalation duration are trivially related because it is impossible to inhale without pausing (unless the speech is ingressive). Nevertheless, Whalen and Kinsella-Shaw (1997) report that pause duration shows more consistent correlations across speakers with respect to sentence length than inhalation duration.

The most interesting link is the one between initial f0 peak and inhalation depth. The literature indicates that f0 declination can be due both to physiological constraints (such as declining subglottal pressure, see Maeda, 1976) and active vocal fold control (Gelfer et al., 1983; Ohala, 1978; Strik, 1994) on the basis of linguistic constraints (Cooper & Sorensen, 1981; Liberman & Pierrehumbert, 1984). This suggests that f0 trends are partly mechanical and partly under the speaker's control (determined by linguistic structure). Mechanisms specific to the initial f0 peak have not been tested extensively. van Heuven (2004) hypothesizes a relationship between initial f0 peak height and inhalation depth. He suggests that ''when the speaker inhales, he has a rough idea of how much material he is going to speak until the next inhalation pause […]. Before longer sentences, then, the speaker will take a deeper breath than before short sentences, so it is the volume of air trapped inside the lungs that primarily determines the high onset pitch rather than some complex computational act the speaker performs on the pitch contour […]'' (van Heuven, 2004, pp. 83–84). In a similar vein, Watson, Ciccia, and Weismer (2003) demonstrate that initiating speech at three very different lung volume levels produced significant change in some of the acoustic variables examined, and, in particular, they find that increasing lung volume at speech initiation leads to higher fundamental frequency values. They offer an explanation with respect to a mechanical link: the inspiratory lowering of the diaphragm pulls on the trachea and the larynx, thereby increasing vocal fold tension.

If initial f0 peak raising is a mere mechanical by-product of inhalation depth, then a strong correlation between the two is expected, i.e., the deeper the inhalation taken before sentence onset, the higher the initial f0 peak. However, given the close link between f0 peak raising and f0 declination, we could also expect that f0 peak raising is not a pure by-product of inhalation depth and is also affected by laryngeal activity. In that case, the correlation will be much weaker.

1.5. Goals of the current study

The goal of the present study is to address the question of whether sentence length and syntactic complexity affect acoustic and respiratory parameters. If so, these parameters can be considered indicators of sentence planning. The relation between the examined parameters will further allow us to gain an insight into the sensitivity of each of the parameters to sentence length and syntactic complexity. This in turn should lead us to a more differentiated understanding of the planning process.

The study focuses on German, a language which has previously been rarely discussed in the context of speech planning and acoustic and respiratory parameters. By examining the effect of the upcoming sentence alone, the pitfalls of some of the previous studies combining sentences preceding and following a pause are avoided. This will allow us to draw more reliable conclusions about planning effects. Furthermore, to the best of our knowledge, the interaction of length and syntactic complexity has not previously been investigated systematically for languages other than English and Swedish. In addition, while this is not the main focus of the study, we plan not just to examine group effects, but also to observe qualitative differences in speaker-specific behaviour, since this forms the necessary substrate for better understanding of flexibility in the speech planning process.

The general predictions derived from the literature review above are:

Hypothesis Ia. Longer sentences will lead to longer pause duration, as speakers will need more time to plan the upcoming sentence.

Hypothesis Ib. Syntactically complex sentences will lead to longer pause duration, as speakers will need more time to plan the complex structure.

Hypothesis II. Longer sentences will lead to higher initial f0 peaks as speakers will need more f0 space to apply declination.

Hypothesis IIIa. Longer sentences will lead to deeper inhalation, since they are produced with longer pause duration which leads to more time for inhalation.

Hypothesis IIIb. Syntactically complex sentences will lead to deeper inhalation.

Hypothesis IV. We expect inhalation duration to be proportional to inhalation depth, and thus to be longer in longer sentences.

These predictions are schematized in Fig. 1.

The relation between syntax and initial f0 peak raising is poorly understood. For initial peak raising, a few studies report no effect of syntax. On the other hand, we cannot exclude a possible effect in German, since syntactic embedding has been found to affect the height of the f0 peaks. Similarly, the effects of syntax on inhalation duration are almost unexplored. As a consequence, no explicit predictions are made for the effects of syntactic complexity on initial f0 peaks and inhalation duration.

The study consists of two experiments. In Experiment I we investigate for all four parameters (pause duration, initial fundamental frequency peak, inhalation depth and inhalation duration) the effects of sentence length (in number of syllables) and syntactic complexity (one clause vs. more clauses). Syntactic complexity is defined in our study in a rough way by the number of clauses. Simpler sentences consist of one clause and syntactically complex sentences of more than one clause. This measure was also used in Whalen and Kinsella-Shaw (1997) and should be seen as a first step towards a better understanding of the influence of syntactic structure on physiological speech planning parameters.

Experiment II is a follow-up study and focuses on the effect of length on the acoustic parameters only. Utterance length is manipulated by modifying the length of the subject constituent (which was kept constant in Experiment I) and the length of the prepositional phrase in subject–verb–prepositional phrase sentences. Experiment I was run on twelve participants and Experiment II on twenty participants.

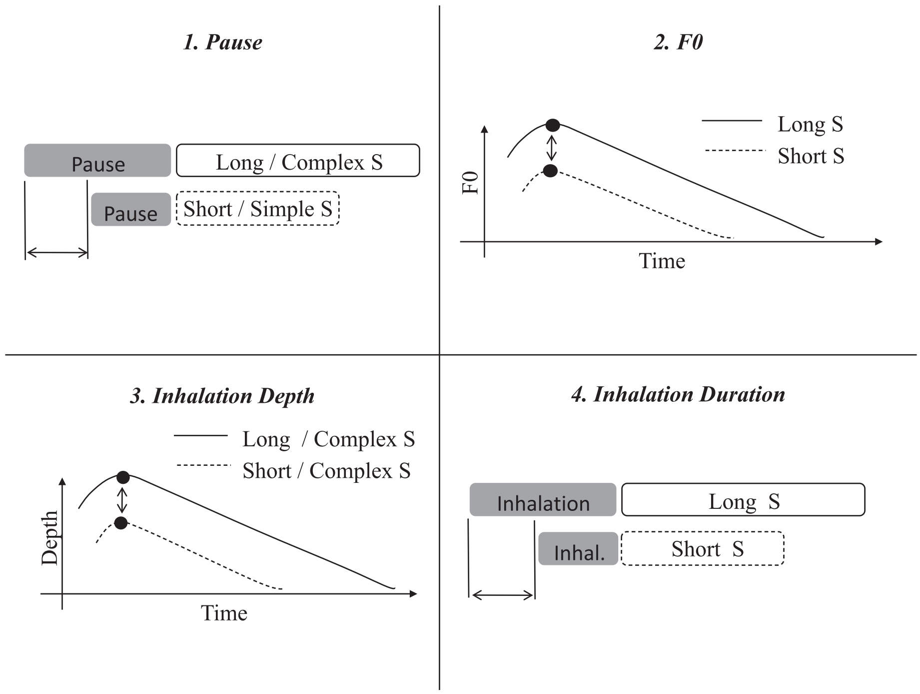

Fig. 1. Schematic illustration of the hypothesized effect of utterance length and syntactic complexity on the four parameters under investigation.

2. Experiment I

2.1. Methods

2.1.1. Stimuli

Before starting the experiment, an acoustic pilot study with ten German speakers was carried out, adapting the methods described in Whalen and Kinsella-Shaw (1997). Since many participants found the task of performing voluntary breathing manoeuvres before starting the target sentence confusing, a different procedure was chosen. The target sentence was always preceded by a long sentence which ended with a colon (*In der Wochenendausgabe der Berliner Zeitung von Anfang Dezember stand geschrieben*: English translation: ‘‘In the weekend issue of the Berlin newspaper from the beginning of December was written:’’). The length of this context sentence was slightly adjusted for each participant so that it would naturally induce the participant to inhale before the target sentence without causing them to be completely out of breath: for participants who spoke faster and were able to realize the context sentence and a short target sentence in one breath, we increased the size of the context sentence by adding ‘‘1996’’ (eight syllables) after ‘‘December’’. The context sentence was kept constant per speaker for the whole experiment.

As opposed to Whalen and Kinsella-Shaw (1997) participants were not instructed to read the target sentence within one breath, but the length of the target sentence was varied in a range that allowed the sentence to be spoken spontaneously within one breath. The corpus consisted of short and long sentences (14 vs. 24 syllables) with two degrees of syntactic complexity, simple and complex (simple: one matrix clause, complex: one matrix clause and at least an additional relative clause). The two factors were fully crossed.

There were two sentences for each condition. All sentences started with one of the two words *Sonja* or *Tonja* (proper names), which are stressed on their first syllable (*Son-* in *Sonja* and *Ton-* in *Tonja*).

Sentences were presented in randomized order, and repeated ten times. In sum, the corpus consisted of 960 observations (8 sentences $\times$ 10 repetitions $\times$ 12 speakers). We included two different words to be able to investigate respiratory effects on different obstruents similarly to Slifka (2003). However, this last point will not be considered here. The target sentences for this experiment are given in Appendix A.

2.1.2. Speakers

Seven females and five males, with no known speech or hearing disorders, participated in this experiment. Their age ranged from 21 to 41 years (mean = 29 years old) Ten speakers came from Berlin or Northern Germany and two speakers from the South. They all spoke Standard German. All speakers were students or academics and naïve with regard to the aim of the experiment. Speakers were paid for their participation.

2.1.3. Procedure

Respiratory and acoustic data were simultaneously recorded. Respiratory activity was monitored by means of Respiratory Inductance Plethysmography (‘RIP’; Respitrace, Ambulatory Monitoring Inc., Ardsley, NY). For this technique the participant wears two elasticized bands each approximately 10 cm from top to bottom edge, one around the rib-cage (top edge roughly at the level of the axilla), and one around the abdomen (top edge roughly at the level of the umbilicus). See e.g., Hixon and Hoit (2005, p. 168) for an illustration. A wire sewn into each elasticized band acts as an expandable loop that modulates the frequency of a small oscillator attached to the body. After demodulation, the resulting signal is proportional to the average cross-sectional area enclosed by each band. Participants were recorded in a straight sitting position. The two respiratory signals were recorded together with the audio signal on a multichannel DAT instrumentation recorder (Sony

PC208x), with low frequency response down to DC. Since for this study we were not interested in the detailed coordination patterns of rib-cage and abdomen the rib-cage and abdomen signals were summed to give a signal proportional to overall lung volume using calibration constants determined by means of an isovolume manoeuvre (see e.g., Hixon, Goldman, & Mead, 1973, especially Fig. 1, for illustration; Augousti, 1997; Konno & Mead, 1967, for further background). Briefly, in this manoeuvre participants repeatedly displace volume between rib-cage and abdomen while keeping the upper airway closed. In such a situation the volume contributions of rib-cage and abdomen can be assumed to be equal and opposite, and the raw voltage signals are weighted appropriately to give this result. Note, also, that since for the purposes of this study we are essentially interested in *relative* volume changes in different linguistic conditions it was not considered necessary to further convert the summed Respitrace signal (expressed in volts) to calibrated absolute values in litres (e.g., using a spirometer).

2.1.4. Measurements

The acoustic analysis was performed by means of PRAAT (Boersma & Weenik, 2011), and included both duration and f0 measurements. Based on visual inspection of the acoustic signal, the onsets and offsets of the pause, the target sentence, and the first word were manually labelled. The pause duration (Parameter I) was calculated as the sum of the last syllable at the end of the context sentence plus the following silent pause. This choice is motivated by the fact that prosodic boundaries affect the duration of the segments close to the boundaries, potentially interacting with the duration of the following silent pause (Ferreira, 1993). The f0 peak (Parameter II) was first automatically detected as the f0 maximum (in Hz) within the first word of the sentence (*Sonja*/*Tonja*). Subsequently, all measurements were checked by visual inspection of the f0 tracks in combination with the waveform and spectrogram for each utterance. When an automatic detection error occurred, the f0 maximum was manually labelled (see Fig. 2 as an example of f0 labelling).

Usually, the f0 maximum occurred on the final (unstressed) syllable of the target word (*-ja*), which is typical of non-final (L*+H) pitch accents in German (Atterer & Ladd, 2004; Braun, 2007; Grabe, 1998; Truckenbrodt, 2002; Uhmann, 1991, *inter alia*). The occurrence of the f0 peak in a sonorant context minimized possible difficulties in peak detection due to segmental perturbations. f0 values were first labelled in Hz and then converted to semitones.

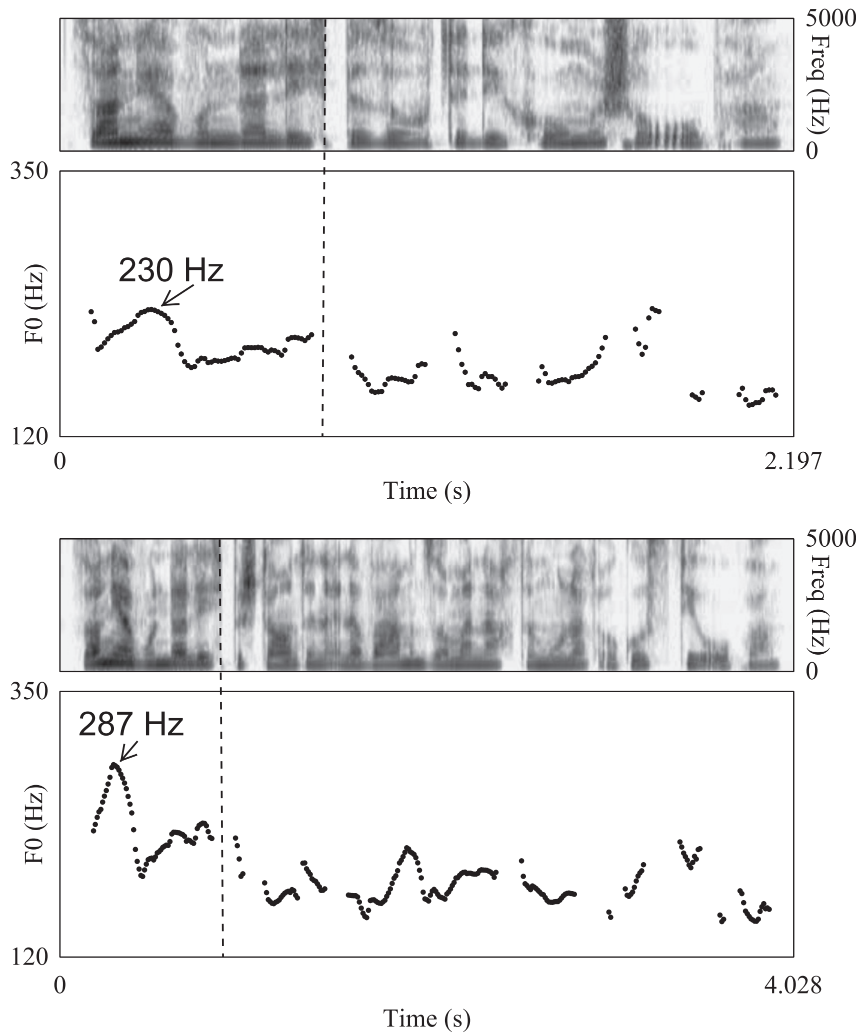

Fig. 2. Example for labelling the f0 peak in the short simplex sentence ''Sonja Wunderlich besuchte die Komische Oper'' (upper panel) and in the long simple sentence ''Sonja Wunderlich bestaunte in einer warmen Sommernacht im Monat August die Oper'' (lower panel). The arrows indicate the temporal location of the utterance-initial f0 peak within the f0 track. The dashed lines mark the end of the Subject constituent (*Sonja Wunderlich*).

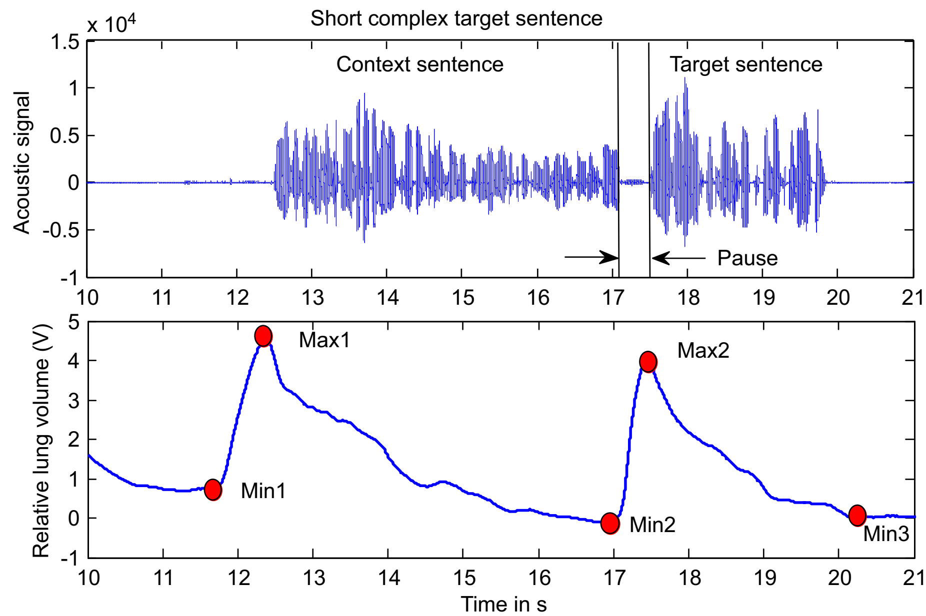

Fig. 3. Example for labelling acoustic and respiratory parameters in a short complex sentence. First track acoustic signal, second track weighted respiratory sum signal of thoracic and abdominal volume. Dots correspond to minima and maxima occurring with respect to the context and the target sentence. Tokens with 3 breathing cycles were not included in the final analysis.

The respiratory parameters were manually labelled on the summed thoracic and abdominal signals by means of a self written graphical interface in MATLAB (The MathWorks, version 2007a) (see Fig. 3). Inhalation minima and maxima (seen as dots in Fig. 3) were labelled for the context and the target sentence. Inhalation depth (Parameter III) occurring before the target sentence was calculated as the difference between the maximal volume expansion at the onset of the target sentence and the minimal volume expansion at the end of the context sentence. Inhalation duration (Parameter IV) was calculated as the temporal difference between the same two data points.

Since our corpus was characterized by a certain amount of variation, some criteria for inclusion of data in the analyses had to be applied to avoid possible statistical biases.

The first criterion was that the target sentence had to contain only a single inhalation peak, realized at the beginning of the sentence, i.e., the target sentence had to be realized in a single breath. Following this criterion, 855 sentences of 960 sentences were retained and 105 sentences were discarded. Specifically, in 82 out of the 105 sentences, a second, additional inhalation peak was realized within the target sentence (see results section on inhalation depth and duration for further details). In 23 out of the 105 sentences (all 23 belonging to the same speaker), no inhalation peaks were realized at all in the target sentence, but a peak was visible only at the beginning of the context sentence, indicating that the context and the target sentences were uttered as a single breath group. The second criterion for data inclusion was that an f0 peak (Parameter II) had to be realized in the utterance-initial word (*Sonja* or *Tonja*). This criterion was applied only for the f0 analysis and led to the exclusion of 298 out of 855 sentences in total. The discarded sentences were characterized by de-accentuation of the target word. One possible explanation for the large number of de-accented target words is that de-accentuation is a strategy to avoid tonal overlap in an accent clash condition. In constituents such as *SONja WUNderlich*, there is only one unstressed syllable (*-ja* in *Sonja*) between the two stressed ones (*SON-* in *Sonja* and *WUN-* in *Wunderlich*). Moreover, the accent on *Sonja* can be delayed and occur after the stressed syllable boundaries (this pattern being typical of L*+H accents in German) so that it can be very close in time with the following one. The suppression of the first pitch accent can thus be a way to ''adjust'' the temporal organization of pitch accents (accentuation patterns will thus be better controlled in Experiment II).

To sum up, the statistical analyses of pause duration, inhalation depth, and inhalation duration are based each on 855 tokens, while the analysis of the initial f0 peak is based on 557 tokens.

2.1.5. Auditory annotation

From informal observations of the acoustic corpus, we noticed that the syntactic and length manipulations also yielded some variation in the prosodic structure of the target sentences. As a consequence, we designed a perceptually-based annotation task to check possible prosodic structure differences across the corpus. The task was performed by three German native speakers: The first author of this paper, a colleague at the ZAS and a student of linguistics at the Humboldt University in Berlin. The student was paid for her participation. The annotators first had to pass a pre-test in which they were trained by way of examples: (a) to find the perceptually most "prominent" words in the sentence (which would allow us to find the location of pitch accents) and (b) to find "perceptual boundaries" between prominent words. The assignments were based on auditory impression only rather than on inspection of acoustic signals. The first annotator transcribed the whole corpus (960 sentences). Since the annotation was very time-consuming, the corpus was split into two sub-groups for the auditory annotation. The second annotator was presented with the first three repetitions of each sentence produced by each speaker, while the third one was presented with the last three repetitions. Thus, the second and the third annotators each had to label 480 sentences. Each utterance was auditorily transcribed by two annotators (the first and the second ones for the first sub-group of the corpus and the first and the third ones for the last sub-group). The sentences were played in a randomized order from a computer through professional headphones in a quiet room at ZAS. The labellers could listen to each sentence as often as they wanted and report their annotation on a printout of the sentences in orthographic form. The annotation task was only partly successful. While the annotators found the task of determining prosodic boundaries generally easy, they all admitted to considerable difficulty in locating prominent words, especially for speakers with a high speech rate and long and/or complex sentences. We thus focussed only on auditory results for the prosodic boundary location. In particular, we were interested in the presence or absence of a boundary after the subject constituent, since the literature on initial f0 peaks had suggested that the prosodic structure and length of the first major constituent in the sentence play a major role in the height of utterance-initial peaks (e.g., Ladd & Johnson, 1987; Prieto et al., 2006), and since prosodic phrasing is known to affect speech planning. We found that both annotators transcribed a prosodic boundary after the subject constituent in 80.2% of the sentences from the whole corpus, and at least one annotator in 95.6% of the sentences. None of the sentences contained a prosodic boundary within the subject. In other words, there was a strong tendency for the subject constituent to be produced as a single prosodic constituent through the whole corpus.

A possible account for this result is that in our experiment subject length worked as an active phrasing constraint: speakers were induced to put a prosodic boundary at the end of the subject constituent because it was always long (containing five syllables and two pitch accents). Our result is in fact in line with findings in some Romance languages (Prieto et al., 2006), in which long subject constituents tended to be phrased as single prosodic constituents.

2.1.6. Statistics

A series of statistical analyses tested separately for the effect of the independent variables on pause duration, utterance-initial f0 peak, inhalation depth and inhalation duration. All the analyses were run by means of the R-environment (R Development Core Team, 2008), and included a series of linear mixed models (Pinheiro & Bates, 2000). Mixed models have emerged as powerful tools for statistical analysis, and have been used in many scientific disciplines including, more recently, phonetics (see, in this journal, Meunier and Espesser (2011) for an example of mixed models run on segmental durations and Calhoun (2012) on f0). Mixed models incorporate both fixed and random factors. Generally, a factor is "fixed" if the aim of the experimenter is to explicitly compare the levels of such a factor, while it is "random" if the levels of the variable are regarded as randomly sampled from a larger population (Baayen, 2008). Moreover, mixed models are preferable to more classical statistical analyses like repeated measure ANOVA, since they can deal with unbalanced data, violation of sphericity assumption and sampling hierarchy (for a comparison between mixed models and repeated measure designs, see Quené and Van den Bergh, 2004). It is still unclear, though, how to calculate the degrees of freedom in models including random factors. Hence the *p*-values are calculated using the method of Monte Carlo sampling by Markov chain (*pMCMC* = Monte Carlo Markov Chain; see Baayen, 2008).

In Experiment I, the following mixed models were used. For pause duration, inhalation depth and inhalation duration the three factors Syntactic complexity (simple/complex), Length (long/short) and Word (*Sonja*/*Tonja*) were treated as fixed factors, whereas Speaker was included as a random factor. For initial f0 peaks, Gender (male/female) was added as a fourth fixed factor, since females have higher fundamental frequency than males (Laver, 1980). Pairwise interactions between the fixed factors were also tested. When interactions were not significant, simpler, additive models were fitted to the data to check the significance of main effects. For the sake of brevity results for pairwise interactions are only reported when they were significant. Otherwise, we only report results for main effects. Because of the complexity of the multiple analyses performed on the dataset, we used an alpha of $pMCMC < 0.01$.

2.2. Results

2.2.1. Parameter I: Pause duration

The mean pause duration was longer before long (564.4 ms) than short (526.4 ms) sentences. On the other hand, pause duration was nearly identical for complex (548 ms) and simple (540.4 ms) sentences. Statistical results revealed a significant effect of Length [$t = -6.1$, $pMCMC < 0.001$] while Syntactic complexity was not significant [$t = 0.05$, $pMCMC = 0.9$]. Word was also significant [$t = 4.8$, $pMCMC < 0.001$]. The Word effect (545 ms pause duration for *Tonja* vs. 523 ms for *Sonja*) can be explained by the fact that the silent closure of /t/ is in our measurements part of the pause (since it cannot be differentiated from a silent pause).

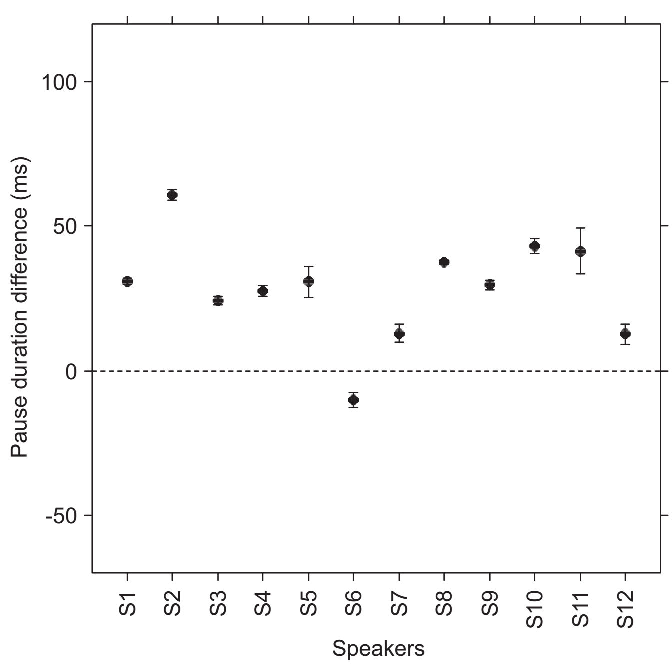

Fig. 4. Means and standard errors for the difference in pause duration between long and short sentences (*y*-axis). Results are split by speakers (*x*-axis). The dashed line at 0 indicates no difference between the two conditions. Positive values indicate that the pauses were longer in long sentences, negative values that they were longer in short sentences. Data are collapsed across syntactic complexity and word type.

Although the group effect of Length was significant, there seems to be some inter-speaker variability. Fig. 4 shows that even though there is an overall tendency for pauses to be longer before longer sentences, the effect of Length is stronger in some speakers (S1, S2, S3, S4, S5, S8, S9, S10, S11) than in others (S6, S7, S12), at least to the extent that can be estimated from this figure. For instance, the mean pause duration difference between long and short sentences was 60 ms in S2 but only 10 ms in S12.

2.2.2. *Parameter II: Initial fundamental frequency peak*

Fig. 5 shows the heights of the f0 peaks in semitones (the reference f0 for semitone calculation is 1 Hz).

The statistical model revealed that neither Length [$t=-0.9$, $pMCMC=0.31$] nor Syntactic complexity were significant [$t=-0.2$, $pMCMC=0.8$]. As expected, Gender [$t=-23.15$, $pMCMC<0.001$] was significant, since the average of the f0 peaks was much lower in male than in female voices. Word was not significant [$t=1.78$, $pMCMC=0.07$], whereas the interaction Gender × Word was significant [$t=-4.06$, $pMCMC<0.001$]. All the other interactions were not significant.

2.2.3. *Parameter III and IV: Inhalation depth and duration*

The average inhalation depth was 3.84 V before long sentences and 3.62 V before short sentences, while inhalation duration was 567.9 ms before long sentences and 523.9 ms before short ones. Differences between syntactically complex and simple sentences were on average much smaller both for inhalation depth (complex: 3.72 V; simple: 3.73 V) and duration (complex: 553.4; simple: 536.2).

Figs. 6 and 7 show an effect of sentence Length on inhalation depth and inhalation duration.

Results are split across speakers (S1–S12). As already found for pause duration, there is a fair amount of speaker-specific behaviour. Nevertheless, for most speakers the difference in inhalation depth is positive, with deeper inhalation for longer sentences, except for S6 and S7. Similarly, most speakers show longer inhalation durations for longer than shorter sentences, except for S6 and S12.

The statistical analysis revealed a significant effect of sentence Length on inhalation depth [$t=-4.3$, $pMCMC<0.001$] and inhalation duration [$t=-3.8$, $pMCMC<0.001$], while Syntactic complexity and Word were not significant for either parameter. While Syntactic complexity had no effects on the depth of inhalation, it had some impact on the number of inhalation peaks within the target sentences. Specifically, we found that in 82 cases (out of the 960 sentences that were collected) there was an additional inhalation peak in the middle of the target sentence. This second inhalation peak occurred in 72 out of these 82 cases in long and complex sentences, in 8 out of the 82 cases in short and complex sentences and only in once out of 82 cases in each of the other two conditions. As discussed in the section Measurements, these sentences were excluded from the statistical analysis because of possible biases due to differences in the number of inhalation peaks. However, such discarded sentences suggest that syntactic complexity leads to the occurrence of additional breaths, but not to a deeper or longer inhalation.

2.2.4. *Relations between the four parameters*

Possible correlations between the acoustic and respiratory parameters were also tested. Specifically, we tested whether pause duration is correlated with inhalation duration, inhalation depth and with the initial f0 peak and whether initial f0 peak is correlated with inhalation depth.

Relationships between acoustic and respiratory parameters were assessed by means of linear mixed models. A significant positive correlation was found between pause duration and inhalation duration [$t=16.7$, $pMCMC<0.001$] and between pause duration and inhalation depth [$t=18$, $pMCMC<0.001$].

Correlations between f0 (calculated again in semitones) and the other parameters were run only for sentences in which a pitch accent was realized in the target word. The correlation between pause duration and f0 peak interacted with sentence length, in that there was a positive correlation between these two parameters only in short sentences [$t=3.42$, $pMCMC<0.001$], but not in long ones [$t=1.44$, $pMCMC=0.07$]. The correlation between initial f0 peak and inhalation depth was not significant [$t=0.21$, $pMCMC=0.83$].

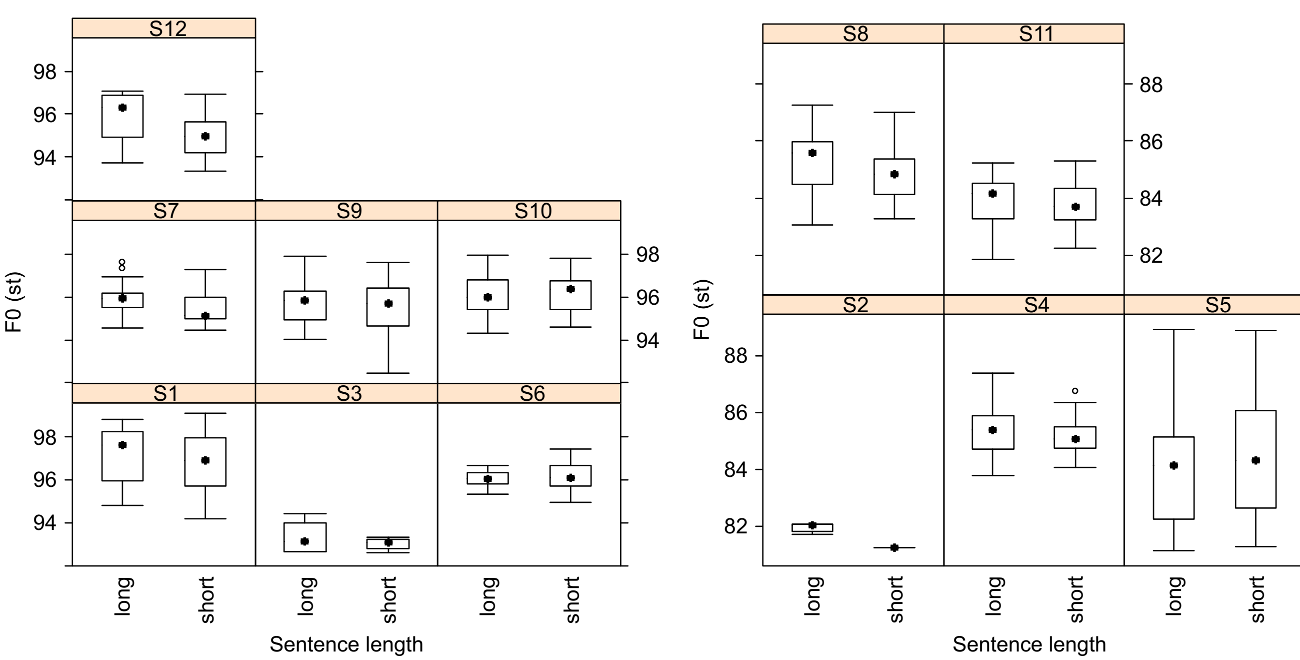

Fig. 5. Boxplots for semitone values of utterance-initial f0 peaks (*y*-axis) against sentence length (*x*-axis) for females (left panel) and males (right panel) and split across speakers. Data are collapsed across syntactic complexity and word type.

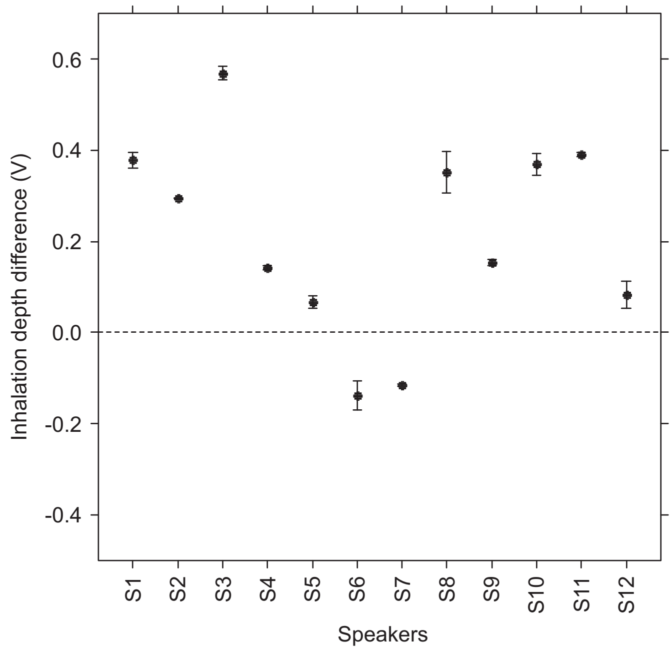

Fig. 6. Means and standard errors for the difference in inhalation depth between long and short sentences (*y*-axis). Results are split by speakers (*x*-axis). The dashed line at 0 indicates no difference between the two conditions. Positive values indicate that inhalation was deeper in long sentences, negative values that it was deeper in short sentences. Data are collapsed across syntactic complexity and word type.

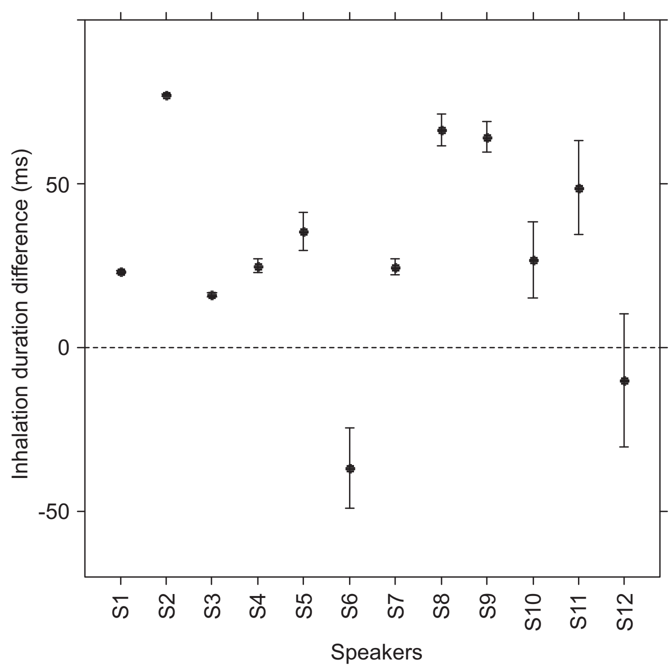

Fig. 7. Means and standard errors for the difference in inhalation duration between long and short sentences (*y*-axis). Results are split by speakers (*x*-axis). The dashed line at 0 indicates no difference between the two conditions. Positive values indicate that inhalation was longer in long sentences, negative values that it was longer in short sentences. Data are collapsed across syntactic complexity and word type.

We were not able to calculate the proportion of variance (R^2) for each of these models, since in mixed models there is no well-established way to calculate the R^2 explained by the fixed factors (and hence the degree of correlation between two variables), and separate it from the variance of the random factor (Speaker). As a consequence, R^2 values are not reported.

3. Discussion

Results for Experiment I show that sentence planning is reflected in the Parameters I, III and IV (pause duration, inhalation depth and inhalation duration).

The effects of sentence Length were highly significant: long sentences led to longer pauses, deeper inhalation and longer inhalation duration than short sentences. The effect of Length on pause duration is in line with findings previously reported in the literature (e.g., Ferreira, 1991; Sternberg et al., 1978; Wheeldon & Lahiri, 1997; Zvonik & Cummins, 2003). We interpret this effect as indicative of a longer planning time needed for the longer sentences as has also been suggested in the cited literature.

The deeper inhalation and the longer inhalation duration for longer compared to shorter sentences are interpreted with respect to a rough anticipation of larger air volume requirements in upcoming long sentences. The results are in line with Whalen and Kinsella-Shaw (1997) and Winkworth et al. (1995). Although results for inhalation depth and duration showed in general significant length effects, the magnitude of such an effect on breathing may often be smaller at the beginning than at the end of the sentence. In fact, Fuchs, Hoole, Vornwald, Gwinner, Velkov, and Krivokapić (2008), in a preliminary analysis of the data presented here, found that in long sentences the thoracic and abdominal volumes reach a considerably lower level at the end of long sentences in comparison to short sentences. Both results can be interpreted as follows: Although speakers plan inhalation depth to some extent, they can also rely on the expiratory reserve volume, i.e., the air which is still in the lungs. So far it is not known whether speakers would show similar breathing patterns if not only one, but several target sentences were to follow. It is possible that participants in our study made particular use of the expiratory reserve volume because they knew that only one target sentence had to be produced and they would be able to take a deep breath after that.

Sentence-initial f0 peaks turned out to show no significant effects, contrary to Hypothesis II. A possible explanation for this result is that prosodic constraints worked against anticipatory effects on the initial f0 rise. It has previously been shown that linguistic structure can affect the scope of planning (see Ferreira, 1991 for syntactic structure, Krivokapić, 2007a, 2007b for prosodic structure). In the current experiment, the size of the subject constituent was kept constant (*Sonja Wunderlich*/*Tonja Wunderlich*). The auditory annotation also revealed that the subject constituent was almost always followed by a prosodic boundary, independent of Length and Syntactic complexity. This means that the sentences had a similar prosodic structure, with the subject constituent being uttered as a single prosodic unit. As a consequence, the prosodic boundary after the subject constituent might have limited the scope of planning for the initial f0 rise to this constituent and might have weakened the effect of Length. If this hypothesis is true, then the utterance-initial f0 peak will be sensitive to Length manipulations only when it is the length of the first constituent of the sentence that is modified. We examine this hypothesis in Experiment II.

One of the reviewers suggested an alternative hypothesis. The lack of f0 effects might be due to the fact that the target sentences were repeated many times so that the subject constituent would be new in the first repetition and ‘‘given’’ in all others. The given status following sentence repetition would have negatively affected the f0 peaks, in the sense that a given first constituent is decreased in f0 as compared to a new one. While we agree that the given status might be encouraged by sentence repetition, external and internal evidence suggest that this hypothesis cannot account for our data. First, experiments in many languages show that initial f0 raising is often found in corpora in which target sentences are repeated many times (see references in the introduction section on f0). Moreover, we will show in Experiment II that in German length changes in the target sentence significantly affect the initial f0 height (despite using repetitions in that experiment), thus suggesting that phonological factors play a major role in determining f0 raising.

The different effects of Length on pause and inhalation duration on the one hand and initial f0 peaks on the other hand may well be due to the fact that these parameters inherently reflect different processes, as suggested in the Introduction. Moreover, the small effects of Length on pause duration and inhalation can be partly due to speaker-specific variation.

The effects of Syntactic complexity were not statistically significant for any of the four parameters investigated (contrary to our Hypotheses Ib and IIIb). This lack of effect is surprising for pause duration, given that syntactic structure often has an effect in non-breathing pauses. It is possible that our short sentences (short simple and short complex) were too short to show an effect of syntactic complexity. In other words, if the sentences are short, they might be so easy for the speaker to prepare that even with added syntactic complexity pause duration will not be affected. For the long sentences on the other hand, our measure of complexity might not be fine-grained enough; while the long complex sentence has two or three clauses, as opposed to one clause in the simple sentences, the long simple sentences have two adjunct phrases, which might add to the syntactic complexity in a different manner. In addition to this, the relative clause sentence in the long sentences modified the object, thus the complexity occurs quite late in the sentence. While overall this adds to the complexity of the sentence, the late relative clause might have only a small effect on pause duration (see Krivokapić, 2010 for a related discussion on the stronger effect of the first prosodic phrase compared to the second).

Syntactic complexity did not show an effect on inhalation depth, in contrast to Hypothesis IIIb. However, it showed an effect in a different measure, inhalation frequency. In complex sentences, the inhalation peaks at the beginning of the target sentence were often followed by an additional peak occurring at a syntactic juncture in syntactically complex sentences.

Finally, the four parameters are only partially correlated. First, pause duration and initial f0 peak are only correlated in short, but not in long sentences. Since the correlations are Length specific, it is impossible to suggest a strong linkage between the two parameters. The lack of relationship between initial f0 peak and inhalation depth argues against van Heuven (2004)’s idea that utterance-initial f0 peaks are the mere by-product of mechanical interactions between respiratory and laryngeal system, and supports the hypothesis that this aspect of f0 production is determined by linguistic structure. Finally, pause duration and inhalation depth as well as inhalation duration are correlated parameters, i.e., the longer the pause, the longer and deeper the inhalation.

4. Experiment II

Experiment II is a shorter follow-up study of Experiment I. It focuses on Parameter I and II, with a particular focus on Parameter II, initial f0 peak. The main goal of Experiment II is to examine whether the initial f0 peak depends on the length of the first subject constituent only, as suggested in Experiment I. We expect higher initial f0 peaks in longer subject constituents. If this is true, the lack of effects for f0 in Experiment I might be due to the fact that the subject constituent was kept constant in length, and sentence length did not affect the initial f0 peak since it was blocked by a prosodic boundary. Furthermore, we tested again, but in a more fine-grained manner, whether the length of the utterance affects pause duration, specifically how the length manipulations of the subject constituent and the prepositional phrase affect pause duration.

4.1. Methods

4.1.1. Stimuli

The stimuli consisted of two sets of sentences composed of subject–verbal–prepositional phrases (S–V–P). The subject constituent and prepositional phrase were proper names that were modified in length for the different conditions. The subject constituent had three length conditions (short = four syllables, medium = five syllables, long = seven syllables) whereas the prepositional phrase had only two (short = two syllables, long = six syllables). The two factors (Subject_length and PP_length) were crossed. The material between them was kept constant and was seven syllables long. Each experimental condition included a subject constituent starting either with the name *Lena* or *Lilli*. Both target words are stressed on the first syllable (*Le-* in *Lena* and *Lil-* in *Lilli*). The choice of target words starting with sonorants allowed us to distinguish the beginning of the sentences from the preceding pauses, thus overcoming the problem we met for /t/ closures in Experiment I. As in Experiment I, we expected a rising pitch accent (L*+H) to be realized on the target words with a delayed peak in the post-accented syllable (*-li* in *Lilli* and *-na* in *Lena*). We expected the difference in vowel identity to have a slight effect on f0, since /ɪ/ and /i/ (in *Lilli*) typically have slightly higher f0 values than /e/ and /a/ (in *Lena*; see Hoole & Mooshammer, 2002 for German). The prepositional phrase was always *aus Suhl* (''from Suhl'') in the short condition and *aus Baden-Würtemberg* (''from Baden-Würtemberg'') in the long condition.

The use of proper names for the subject and prepositional phrase was aimed at keeping the syntactic and semantic structure of the sentences constant across the length manipulations. Moreover, there were always two unaccented syllables between the first two pitch accents in the sentence (e.g., *LILli-MarLEN* vs. *LILli-MaTHILda*) in order to avoid possible tonal clash effects.

All target sentences were preceded by the same context sentence (*Auf dem Zettel steht geschrieben*: ''On the paper it was written:'') ending with a colon in order to induce the speakers to produce a pause between the two sentences. Sentences were presented in randomized order and interspersed among other sentences from two unrelated studies. Each sentence was presented five times to twenty speakers. A full list of the target sentences is given in Appendix B. In sum, the corpus consisted of 1200 observations (3 subject constituent length variations × 2 prepositional phrase length variations × 2 target words × 5 repetitions × 20 speakers).

4.1.2. Speakers

Eleven females and nine males, with no known speech, language or hearing disorders, participated in this experiment. Fifteen of them came from North-East Germany while five came from South-Germany. They all spoke Standard German. The speakers' age ranged between 17 and 38 years old (average: 27 years old). All speakers were students or academics and naïve with regard to the aim of the experiment. Speakers were paid for their participation.

4.1.3. Procedure

The stimuli were recorded in an anechoic room at ZAS with a sampling rate of 24 kHz. Context and target sentences were presented on a computer screen printed each in one line in order to ensure that participants did not place prosodic boundaries due to line breaks. Speakers were asked to read the corpus in a natural way, but were not given any explicit instructions about the prosody to be employed. Speakers were able to repeat a sentence when they felt they had read it in an unnatural way, or in case of disfluencies. The acoustic recording lasted around 35 min.

4.1.4. Measurements

Acoustic data were analysed using PRAAT. Labelling procedures for pause duration and initial f0 peaks were the same as the ones employed for Experiment I.

In 74 out of 1200 sentences (6.1%), it was impossible to detect silent pauses since the segmental region between the context and the target sentences was characterized by glottalization (acoustically manifested mainly by the presence of irregular f0 periods and decrease in amplitude) which made it very difficult to place the boundaries between the two sentences. In fact, glottalization is another phonetic marker of prosodic edges in German, and it can occur even in the absence of silent pauses (Peters, 2005). These sentences were discarded from the statistical analysis of pause duration. For pause duration, 1126 observations were fit in the analysis.

As expected, an utterance-initial f0 peak was in most cases realized on the first word of the compound names (*Lilli* or *Lena*) while only 40 sentences out of 1200 (3.3%) did not match this criterion. Such sentences were discarded from further f0 analysis, yielding 1160 observations.

4.1.5. Statistics

Linear mixed models were run separately on pause duration and on the height of utterance-initial f0 peaks. For pause duration, the Subject_length (short/medium/long), the PP_length (short/long) and Word (*Lilli*/*Lena*) were included as fixed factors. For the initial f0 peak analysis, Gender (male/female) was also included as the fourth fixed factor. Speakers always constituted the random factor. Pairwise interactions between factors were also calculated, but they were factored out from the models when not significant. The cut-off point for significance was $pMCMC < 0.01$.

4.2. Results

4.2.1. Parameter I: Pause duration

Pause duration showed no effects of Subject_length or PP_length: the average pause duration was 350 ms when followed by a long subject constituent, 349.5 ms when followed by a medium, and 345.9 ms when followed by a short subject constituent. Pauses were 351.5 ms when the sentence contained a long and 345 ms when it contained a short prepositional phrase. Neither of the length manipulations had a significant effect on pause duration.

Fig. 8 shows the difference in pause duration between long/medium and medium/short subject length across the 20 speakers. On the *y*-axis, positive values indicate that pause duration was longer before longer subject constituents, while negative values indicate that pause duration was longer before shorter subject constituents. The figure indicates high variability across the subject constituent manipulations. Moreover, it also displays speaker-specific behaviour for this parameter.

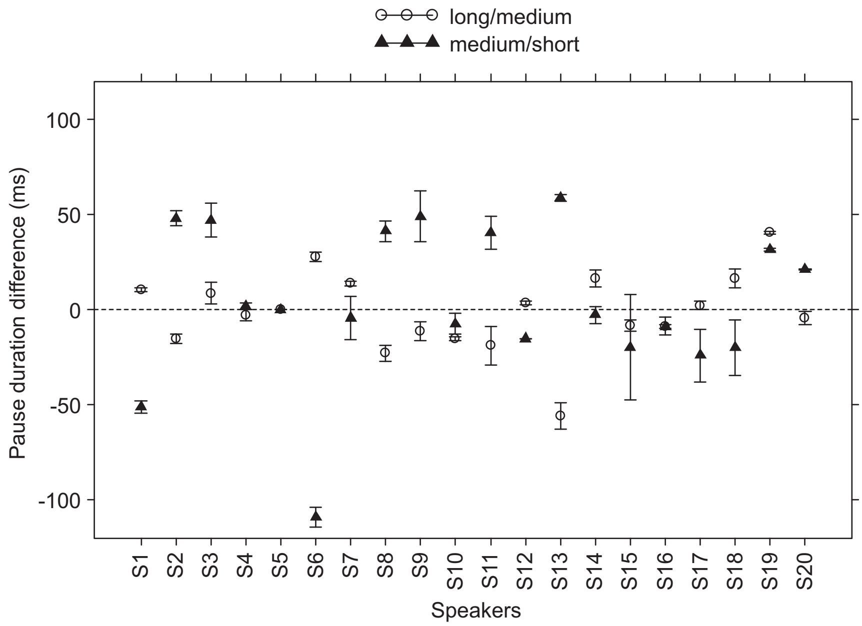

Fig. 8. Means and standard errors for the difference in pause duration between long and medium Subjects (unfilled dots) and medium and short Subjects (filled triangles). Results are split by speakers (*x*-axis). The dashed line at 0 indicates no difference between the two conditions. Positive values indicate that the pauses were longer in the first element of the pairwise comparisons, negative values that they were longer in the second element. Data are collapsed across Prepositional phrase length and Word type.

4.2.2. Parameter II: Initial f0 peak

The height of the utterance-initial f0 peak was more sensitive to the length manipulations of the subject constituent than to manipulations of the prepositional phrase. On average, the initial f0 peak was 90.4 st (197.2 Hz) in long, 90 st (192.5 Hz) in medium and 89.7 st (189.5 Hz) in short subject constituents, i.e., it was 0.7 semitones (7.7 Hz) higher in long than in short subject constituents, respectively. Long prepositional phrases led to an increase of 0.15 semitones (1.3 Hz) compared to short ones.

The magnitude of the effect of Subject_length was quite variable across speakers (see Fig. 9), but 14 of the 20 speakers showed values consistent with the overall trend for higher initial f0 peaks as Subject_length increased (females: S2, S3, S4, S5, S9, S12, S16, S19; males: S1, S10, S13, S17, S18, S20). The greatest f0 difference between long and short subject constituents amounted to 1.45 semitones (21 Hz) for speaker S4. Longer prepositional phrases were realized with higher initial f0 peak for six participants (S8, S10, S13, S17, S18, S20 with semitone values between 0.26 and 0.87 (3.84 and 6.65 Hz respectively)). The other fourteen participants produced only marginal differences which were close to f0 measurement error.

The statistical analysis showed that contrasts among the subject constituent levels were significant (long vs. medium: [$t=-5.73$; $pMCMC<0.001$], long vs. short [$t=-9.46$; $pMCMC<0.001$], medium vs. short: [$t=-4.04$; $pMCMC<0.001$]). The contrast between long and short prepositional phrases across the subject constituent manipulations was in fact also significant [$t=-2.92$, $pMCMC<0.01$]. However, we doubt that this result has any functional relevance: as noted above long prepositional phrases were a mere 0.15 st higher. Interactions between Subject_length and PP_length were not significant. As expected, Gender turned out to be a significant factor [$t=-9.58$, $pMCMC<0.001$]. Word also showed the expected f0 pattern such that the f0 peak was higher in *Lilli* than in *Lena* [$t=9.12$; $pMCMC<0.001$].

5. Discussion

The results for pause duration in Experiment II were somewhat unexpected, since we did not find any effect of length variations on pause duration, contrary to previous studies. One possible explanation for the lack of the effect of length is that the stimuli in this experiment were too similar to each other, and that, despite the filler sentences, this similarity obscured the effects of planning, i.e., made the planning process too easy. Specifically, each sentence varied in the length of the subject constituent and in the length of the prepositional phrase, but the two prepositional phrases differed only in the noun phrase (*aus Suhl* vs. *aus Baden-Würtemberg*). It is possible that speakers became aware of this structure and planned most of the prepositional phrase during the preceding sequence *ist eine berühmte Frau* (‘‘is a famous woman’’), which was always the same across the target sentences. This would mean that speakers produced the sentences in a highly incremental fashion. If this assumption is correct, then mostly the effects of the subject constituent would be noticeable in the measures we used (since the subject phrase was the first constituent). In the case of pauses, the effect of the first constituent then seems too weak to be captured.

Experiment II showed a clear effect of the Subject_length on the utterance initial f0 peak. The mean difference between short and long subject constituents (7.7 Hz) is 3.9%, which is of the order of the just noticeable difference in f0 perception (following a review of the literature Knight & Nolan, 2006, for example, use a figure of 4% for their 2006 investigation). The utterance initial f0 peak is thus sensitive to even small length changes in the first constituent, since the difference between short and long subject constituents was only three syllables.

The effect of the length of the prepositional phrase on the initial f0 peak, while significant, appears too small (1.3 Hz) to be functionally relevant.

These results are in line with Ladd and Johnson (1987) and Prieto et al. (2006), but our study confirms the earlier findings for a substantially larger dataset. In particular, we confirm Ladd and Johnson (1987)'s claim that the height of the first f0 peak in a sentence is related to the length/

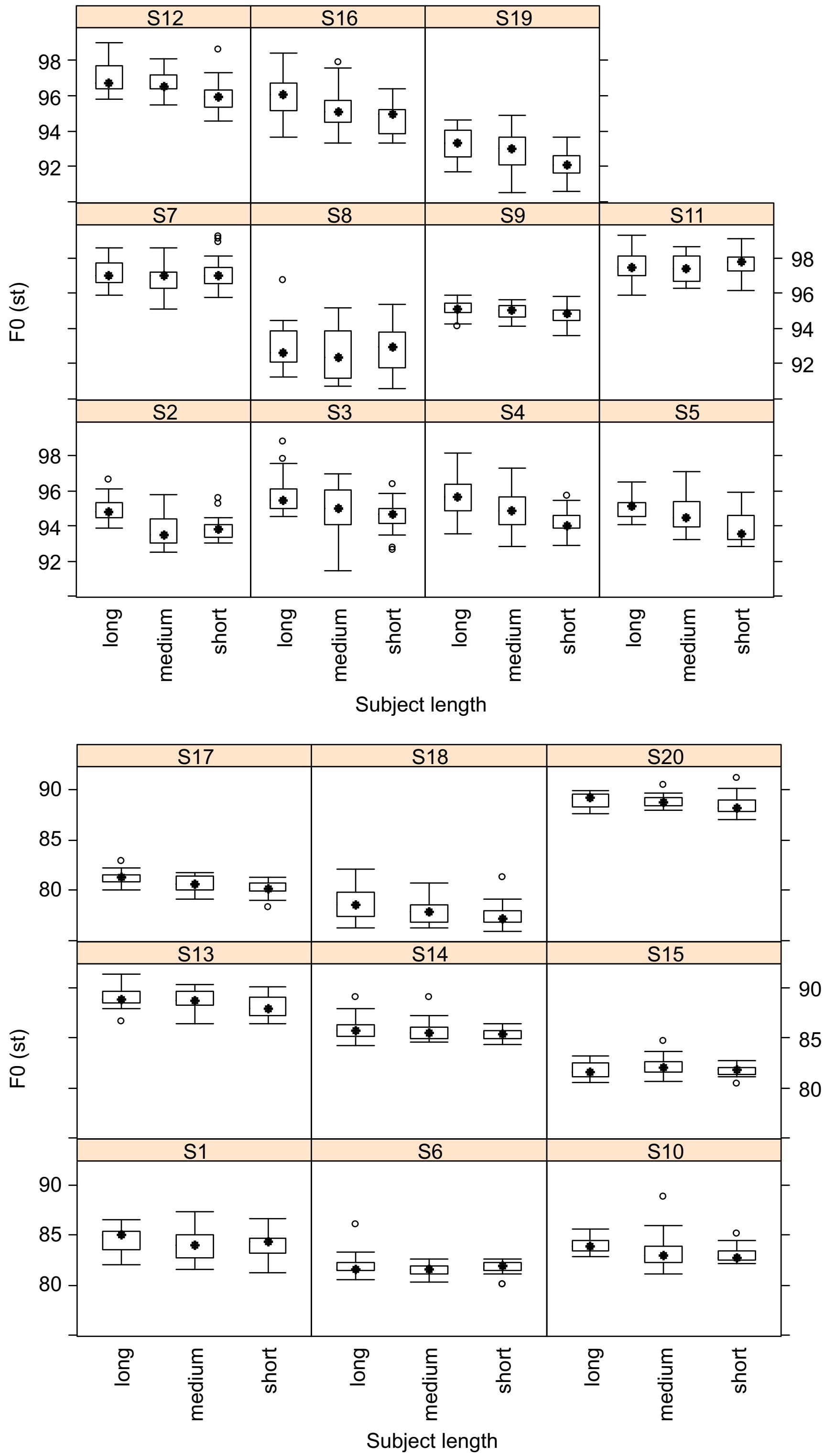

Fig. 9. Boxplots for semitone values of utterance-initial f0 peaks (*y*-axis) against Subject_length (*x*-axis), for females (upper panel) and males (lower panel) and split across speakers. Data are collapsed across Prepositional phrase length and Word type.

prosodic complexity of the first constituent, and not to the length of the whole sentence. Here, the non-significant results for the initial f0 peak in Experiment I and the strong effects in Experiment II indicate that the length of the first constituent is a key factor.

In other words, the height of the initial f0 peak is a parameter showing the planning of the first constituent and is consequently sensitive to the distribution of prosodic boundaries. This suggests that the f0 contour is mainly planned in an incremental way, with the f0 peak height being affected by the length of the constituent in which it is realized. The incrementality of the f0 planning would in fact explain why the initial f0 peak is strongly affected by length manipulations of the first constituent (the Subject) but much less by length manipulations of other constituents in the sentence (e.g., the prepositional phrase).

Finally, our study provides further indications for speaker-specific behaviour (as shown in Fig. 9), with most but not all speakers showing an effect of Subject_length, thus supporting the idea that utterance-initial f0 peak raising is an optional mechanism, in line with the 'soft' planning view. This also suggests the existence of planning strategies by which the scope of the planning unit might be flexibly adapted by the speakers.

6. Summary and conclusion

We investigated speech planning processes in German in two experiments. The goal was to examine how parameters of different aspects of planning are affected by sentence length and syntactic complexity.

The first experiment showed an effect of Length such that longer sentences lead to longer pause duration, longer inhalation duration, and deeper inhalation. We did not observe an effect of syntactic structure on the four measured parameters (but we did find an effect on inhalation frequency), and there was no effect of either Length or Syntactic complexity on the initial f0 peak. It was assumed that the lack of any effect on the initial f0 peak might be due to the occurrence of a prosodic boundary after the first subject constituent. A lack of an effect of syntactic complexity was less surprising with respect to inhalation depth and inhalation duration and it is in line with previous findings by Whalen and Kinsella-Shaw (1997). However, for pause duration our results for syntactic complexity are in contrast to previous studies. We suppose that our definition of syntactic complexity as the number of clauses was too rough and a more fine-grained syntactic analysis is required.

Comparing different parameters with each other, it turned out that f0 is not correlated with inhalation depth. This result speaks against the idea that f0 height is the by-product of a mechanical interaction between the respiratory and the laryngeal system, thus supporting the idea that f0 is actively adjusted by the speaker according to the linguistic structure.

In Experiment II the Length of the first subject constituent and the prepositional phrase were varied to test whether the initial f0 peak is only sensitive to the first constituent. Results from Experiment II confirm this hypothesis. The lack of effect of length manipulations on pause duration was argued to be due to the overall small variations in length and due to task constraints. However, this lack of effect is surprising and merits further research.

The fact that in Experiment I the length of the sentence had an effect on pause duration, inhalation duration, and inhalation depth but not on f0, and that we only found an effect of the length of the first constituent on f0 in Experiment II supports the idea that these parameters reflect different scopes of planning. We suggest that the initial f0 peak is a more *local parameter of planning*, while pause duration, inhalation duration, and inhalation depth are more *global parameters of planning*. F0 also seems to be the parameter most sensitive to small changes, as it was only f0 and not pause duration that showed an effect of the length manipulations in Experiment II.

Finally, we also find support for the suggestion that the effects of the independent variables on each of the parameters are not only driven by linguistic constraints, but are also speaker-specific. The psycholinguistic literature offers some explanations related to how cognitive differences among speakers might affect the speech planning process. For instance, Swets et al. (2007) argued that there is an effect of working memory (i.e., the memory sub-system which is responsible for the active maintenance of action plans or goal states in face of ongoing processing and/or distractions) on prosodic structure, such that readers with a low working memory span are more likely to chunk a text into smaller prosodic phrases. Similarly, Ferreira and Swets (2002) and Wagner et al. (2010) found that an increase of cognitive load leads speakers to decrease the scope of planning, thus providing evidence that the scope of planning is flexible. If cognitive differences can (at least partly) account for speaker-specific differences in the acoustic and respiratory parameters, this indicates that looking at speaker-specific behaviour would be as important as looking at effects of linguistic factors since speaker-specific behaviour could help to shed light on issues concerning speech planning. However, further research needs to be carried out in order to clarify more precisely the effects of speaker characteristics on utterance planning.

Acknowledgements

This work was supported by a grant from the German Ministry for Education and Research to PB1 at ZAS (01UG0711) and the French–German University (DFH) Saarbrücken to the PILIOS project. We also thank Jörg Dreyer for technical support, Pascal Perrier for comments on an earlier draft, Matthias Ziervogel, Anna Sapronova and Susanne Drechsel for their help in labelling the data, and the editor and three anonymous reviewers for helpful suggestions. This work is dedicated to Dieter Fuchs and Giovanni Petrone.

Appendix A. Stimuli for Experiment I

short (14 syllables) and syntactically simple (one clause) sentences
Sonja Wunderlich besuchte die Komische Oper.
(Sonja Wunderlich visited the Comic Opera.)
Tonja Wunderlich besuchte die Komische Oper.
(Tonja Wunderlich visited the Comic Opera.)

short (14 syllables) and syntactically complex (two clauses) sentences
Sonja Wunderlich, die Tanz liebt, besuchte die Oper.
(Sonja Wunderlich, who loves dance, visited the Opera.)
Tonja Wunderlich, der Tanz liebt, besuchte die Oper.
(Tonja Wunderlich, who loves dance, visited the Opera.)

long (24 syllables) and syntactically simple (one clause) sentences
Sonja Wunderlich bestaunte in einer warmen Sommernacht im Monat August die Oper.
(Sonja Wunderlich marveled at the opera during a warm summer night in August.)

Tonja Wunderlich bestaunte in einer warmen Sommernacht im Monat August die Oper.
(Tonja Wunderlich marveled at the opera during a warm summer night in August.)

long (24 syllables) and syntactically complex (3 clauses) sentences
Sonja Wunderlich sagte einem Freund, der uns morgens anrief, sie begeistert sich für Oper.
(Sonja Wunderlich told a friend who phoned us in the morning that she is crazy about opera.)
Tonja Wunderlich sagte einem Freund, der uns morgens anrief, er begeistert sich für Oper.
(Tonja Wunderlich told a friend who phoned us in the morning that he is crazy about opera.)

Appendix B. Stimuli for Experiment II

short subject (four syllables) and short prepositional phrase (nine syllables) = 13 syllables
Lilli-Marlen ist eine berühmte Frau aus Suhl.
(Lilli Marlen is a famous woman from Suhl.)
Lena-Marie ist eine berühmte Frau aus Suhl.
(Lena-Marie is a famous woman from Suhl.)

medium subject (five syllables) and short prepositional phrase (nine syllables) = 14 syllables
Lilli-Matthilda ist eine berühmte Frau aus Suhl.
(Lilli Matthilda is a famous woman from Suhl.)
Lena-Johanna ist eine berühmte Frau aus Suhl.
(Lena-Johanna is a famous woman from Suhl.)

long subject (seven syllables) and short prepositional phrase (nine syllables) = 16 syllables
Lilli-Matthilda Müller ist eine berühmte Frau aus Suhl.
(Lilli Matthilda Müller is a famous woman from Suhl.)
Lena-Johanna Bayer ist eine berühmte Frau aus Suhl.
(Lena Johanna Bayer is a famous woman from Suhl.)

short subject (four syllables) and long prepositional phrase (11 syllables) = 17 syllables
Lilli-Marlen ist eine berühmte Frau aus Baden-Würtemberg.
(Lilli Marlen is a famous woman from Baden-Würtemberg.)
Lena-Marie ist eine berühmte Frau aus Baden-Würtemberg.
(Lena-Marie is a famous woman from Baden-Würtemberg.)

medium subject (five syllables) and long prepositional phrase (11 syllables) = 18 syllables
Lilli-Matthilda ist eine berühmte Frau aus Baden-Würtemberg.
(Lilli Matthilda is a famous woman from Baden-Würtemberg.)
Lena-Johanna ist eine berühmte Frau aus Baden-Würtemberg.
(Lena-Johanna is a famous woman from Baden-Würtemberg.)

long subject (seven syllables) and long prepositional phrase (11 syllables) = 20 syllables
Lilli-Matthilda Müller ist eine berühmte Frau aus Baden-Würtemberg.
(Lilli-Matthilda Müller is a famous woman from Baden-Würtemberg.)
Lena-Johanna Bayer ist eine berühmte Frau aus Baden-Würtemberg.
(Lena-Johanna Bayer is a famous woman from Baden-Würtemberg.)

B.1. Filler sentences

Opa muss heute leider zu Hause bleiben.
(Grandpa has to stay at home.)
Oper mögen besonders alte Menschen gern.
(Old people like opera particularly.)
Am anderen Ende der Stadt steht Reginas Oper.
(Regina's Opera is located at the other end of the city.)
Übermorgen Nachmittag treffe ich Mamas Opa.
(After tomorrow afternoon I will meet mother's grandpa.)
Die Krankenschwester meint, dass Opa bestimmt Kinder mag.
(The nurse says that grandpa probably likes children.)
Der Dirigent hat gesagt, dass Oper bezaubernd ist.
(The conductor said opera is fascinating.)
Die Veranstaltung muss wegen Nässe in Wasser fallen.
(The event was rained off.)

Am Montagmittag wird Wäsche gewaschen und gebügelt.
(On Monday noon laundry will be done and ironed.)

References

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of ''segmental anchoring'' of F0: Evidence from German. *Journal of Phonetics*, *32*(2), 177–197.
Augousti, A. (1997). A theoretical study of the robustness of the isovolume calibration method for a two-compartment model of breathing, based on an analysis of the connected cylinders model. *Physics in Medicine & Biology*, *42*, 283–291.
Avesani, C. (1987). Declination and sentence intonation in Italian. In: P. M. Bertinetto (Ed.), *Quanderni del laboratorio di linguistica della Scuola Normale Superiore di Pisa* (pp. 8–23).
Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Boersma, P., & Weenik, D. (2011). *Praat: Doing phonetics by computer [Computer program]*. Version 5.3. Retrieved from 〈http://www.praat.org/〉.
Braun, B. (2007). Effects of dialect and context in the realisation of German prenuclear accents. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 961–964). Saarbrücken, Germany.
Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, *15.*
Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, *40*, 329–349.
Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition*, *84*, 73–111.
Cohen, A., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua*, *19*, 177–192.
Connell, B. (2003). Pitch realization and the four tones of Mambila. In: S. Kaji (Ed.), *Cross-linguistic studies of tonal phenomena* (pp. 181–197). Tokyo: Research Institute for the Languages and Cultures of Asia and Africa.
Connell, B. (2004). Tone, utterance length and F0 scaling. In *Proceedings of the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, 28–31 March 2004*. Beijing, China: ISCA.
Conrad, B., Thalacker, S., & Schönle, P. (1983). Speech respiration as an indicator of integrative contextual processing. *Folia Phoniatrica*, *35*, 220–225.
Cooper, W. E., & Sorensen, J. (1981). *Fundamental frequency in sentence production*. Springer Verlag.
Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
Core-Team, R.-D. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS*, *6*(5), e19792.
Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147.
Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233.
Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*, 233–253.
Ferreira, V. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, *35*, 724–755.
Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*, 1151–1177.
Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.
Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*(4), 680–703.
Féry, C., & Kentner, G. (2010). The prosody of embedded coordinations in German and Hindi. In *Proceedings of the Speech Prosody*. Chicago. Retrieved from 〈http://speechprosody2010.illinois.edu/papers/100014.pdf〉.
Féry, C., & Truckenbrodt, H. (2005). Sisterhood and tonal scaling. *Studia Linguistica, Special Issue: Boundaries in Intonational Phonology*, *59*(2/3), 223–243.
Fuchs, S., Hoole, P., Vornwald, D., Gwinner, A., Velkov, H., & Krivokapić, J. (2008). The control of speech breathing in relation to the upcoming sentence. In *Proceedings of the 8th International Seminar on Speech Production* (pp. 77–80). Strasbourg.
Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, *22*, 477–492.
Gayraud, F., & Martinie, B. (2008). Does structural complexity necessarily imply processing difficulty? *Journal of Psycholinguistic Research*, *37*, 21–31.
Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458.
Gelfer, C., Harris, K., Collier, R., & Baer, T. (1983). Is declination actively controlled? In: I. R. Titze, & C. Scherer (Eds.), *Vocal fold physiology* (pp. 113–125). Denver, Colorado: The Denver Center for the Performing Arts.
Goldman-Eisler, F. (1968). *Psycholinguistics. Experiments in spontaneous speech*. London and New York: Academic Press.
Grabe, E. (1998). *Comparative intonational phonology: English and German*. Ph.D. Dissertation. Radboud University Nijmegen.
Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica*, *36*(2), 98–114.
Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*, 51–58.
Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). Temporal patterns of cognitive activity and breath control in speech. *Language and Speech*, *8*, 236–242.
Hird, K., & Kirsner, K. (2002). The relationship between prosody and breathing in spontaneous discourse. *Brain and Language*, *80*(3), 536–555.
Hixon, T., & Hoit, J. (2005). *Evaluation and management of speech breathing disorders: Principles and methods*. Redington Brown.
Hixon, T. J., Goldman, M. D., & Mead, J. (1973). Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. *Journal of Speech and Hearing Research*, *16*, 78–115.
Hoit, J. D., Solomon, N. P., & Hixon, T. J. (1993). Effect of lung volume on Voice Onset Time (VOT). *Journal of Speech, Language and Hearing Research*, *36*, 516–520.
Hoole, P., & Mooshammer, C. (2002). Articulatory analysis of the German vowel system. In: P. Auer, P. Gilles, & H. Spiekermann (Eds.), *Silbenschnitt und Tonakzente* (pp. 129–152). Tübingen: Niemeyer.
Hoole, P., & Ziegler, W. (1997). A comparison of normals' and aphasics' ability to plan respiratory activity in overt and covert speech. In: W. Hulstijn, H. Peters, & P. van Lieshout (Eds.), *Speech production: Motor control, brain research and fluency disorders* (pp. 205–212). Amsterdam: Elsevier.
Huber, J., & Spruill, J. (2008). Age-related changes to speech breathing with increased vocal loudness. *Journal of Speech, Language, and Hearing Research*, *51*, 651–668.
Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, *101*, 112–156.
Kentner, G. (2007). Length, ordering preference and intonational phrasing: Evidence from pauses. In *Proceedings of the Interspeech* (pp. 2637–2640). Antwerp.
Knight, R.-A., & Nolan, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association*, *36*(1), 21–38.
Konno, K., & Mead, J. (1967). Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology*, *22*(3), 407–422.
Krivokapić, J. (2007a). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*, 162–179.
Krivokapić, J. (2007b). *The planning, production, and perception of prosodic structure*. Ph.D. Thesis. University of Southern California.
Krivokapić, J. (2010). Speech planning and phrase length. In *Proceedings of the Speech Prosody* (paper number 100311). Chicago.
Krivokapić, J. (2012). Prosodic planning in speech production. In: S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 157–190). Peter Lang.
Ladd, D. R., & Johnson, C. (1987). Metrical factors in the scaling of sentence-initial accent peaks. *Phonetica*, *44*, 238–245.
Laniran, Y., & Clements, G. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*(2), 203–250.
Lashley, K. S. (1951). The problem of serial order in behaviour. In *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley and Sons.
Latash, M. L., & Mikaelin, I. L. (2011). A logarithmic speed-difficulty trade-off in speech production. *Motor Control*, *15*, 52–67.
Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In: M. Aronoff, & R. T. Oehrle (Eds.), *Language sound structure* (pp. 157–234). Cambridge: MIT Press.
Maeda, S. (1976). *A characterization of American English intonation*. Ph.D. Thesis. MIT.
Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C. L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*, 177–192.
McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, *44*, 128–143.
McFarland, D. H., & Smith, A. (1992). Effects of vocal task and respiratory phase on prephonatory chest wall movements. *Journal of Speech, Language and Hearing Research*, *35*, 971–982.
Meyer, A. S. (2004). The use of eye tracking in studies of sentence generation. In: J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 191–212). Hove: Psychology Press.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, *57*, 210–231.
Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271–278.
Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech, Language and Hearing Research*, *39*(1), 93–104.
Ohala, J. (1978). Production of tone. In: V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.
Peters, B. (2005). Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, *35a*, 203–345.
Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Ph.D. Thesis. MIT.
Pierrehumbert, J., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-Plus. Statistics and computing series*. New York, NY: Springer Verlag.
Prieto, P., D'Imperio, M., Elordieta, G., Frota, S. & Vigário, M. (2006). Evidence for soft preplanning in tonal production: Initial scaling in Romance. In *Proceedings of Speech Prosody* (pp. 803–806). Dresden: TUD Press Verlag der Wissenschaften GmbH.
Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*, 445–473.
Quené, H., & Van den Bergh, H. (2004). On multi-level modelling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*, 103–121.
Rialland, A. (2001). Anticipatory raising in downstep realization: Evidence for preplanning in tone production. In: S. Kaji (Ed.), *Proceedings of symposium cross-linguistic studies of tonal phenomena: Tonogenesis, Japanese accentology, and other topics* (pp. 301–321). Tokyo: University of Foreign Studies.
Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–participle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904–921.
Russell, N., & Stathopoulos, E. (1998). Lung volume changes in children and adults during speech production. *Journal of Speech and Hearing Research*, *31*, 146–155.
Sanderman, A.A. & Collier, R. (1995). Prosodic phrasing at the sentence level. In *Producing speech: Contemporary issues. For Katherine Safford Harris* (pp. 321–332). New York: American Institute of Physics.
Slifka, J. (2003). Respiratory constraints on speech production: Starting an utterance. *Journal of the Acoustical Society of America*, *114*(6), 3343–3353.
Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, *20*(2), 171–186.
Sternberg, S., Monsell, S., Knoll, R.L. & Wright, C.E. (1978). *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.
Strangert, E. (1991). Pausing in texts read aloud. In *Proceedings of the XIIth International Congress of Phonetic Sciences* (pp. 238–241). Aix-en-Provence, France: Université de Provence, Service des Publications.
Strangert, E. (1997). Relating prosody to syntax: Boundary signaling in Swedish. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.), *Proceedings of the 5th European Conference on Speech Communication and Technology* (pp. 239–242).
Strik, H. (1994). *Physiological control and behaviour of the voice source in the production of prosody.* Ph.D. Thesis. Catholic University Nijmegen.
Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experiment Psychology: General*, *136*, 64–81.
Terken, J. & Collier, R. (1992). Syntactic influences on prosody. In *Speech perception, production and linguistic structure* (pp. 427–438). Amsterdam, Washington, Oxford: IOS Press and Tokyo, Osaka, Kyoto: Ohmsha.
't Hart, J. (1979). Explorations in automatic stylization of F0 curves. *IPO Annual Progress Report*, *14*, 61–65.
Thorsen, N. (1986). Sentence intonation in textual context—Supplementary data. *Journal of the Acoustical Society of America*, *80*(4), 1041–1047.
Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology*, *19*, 77–120.
Uhmann, S. (1991). *Fokusphonologie*. Tübingen: Niemeyer.
Van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992). Downstep in Dutch: Implications for a model. In: G. Docherty, & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gestures, segment, prosody* (pp. 335–359). Cambridge: Cambridge University Press.
van Heuven, V. J. (2004). Planning in speech melody: Production and perception of downstep in Dutch. In: H. Quené, & V. J. van Heuven (Eds.), *On speech and languages: Studies for Sieb G. Nooteboom* (pp. 83–93). Utrecht: Netherlands Graduate School of Linguistics (LOT).
Vanrell Bosch, M. (2007). Effect of utterance length on F0 scaling. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1165–1168) Saarbrücken.
Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.
Watson, P. J., Ciccia, A. H., & Weismer, G. (2003). The relation of lung volume initiation to selected acoustic properties of speech. *Journal of the Acoustical Society of America*, *13*(5), 2812–2819.
Whalen, D. H., & Kinsella-Shaw, J. M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, *54*(3–4), 138–152.
Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381.
Winkworth, A. A., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity and linguistic factors. *Journal of Speech and Hearing Research*, *37*, 535–556.
Winkworth, A. L., Davis, P. J., Adams, R. D., & Ellis, E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech, Language and Hearing Research*, *38*, 124–144.
Zvonik, E., & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In *Proceedings of the Eurospeech* (pp. 777–780). Geneva, Switzerland.