Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands

Susanne Fuchs, Caterina Petrone, Amélie Rochet-Capellan, Uwe D. Reichel, Laura L. Koenig

To cite this version:

HAL Id: hal-01164773
https://hal.science/hal-01164773

Submitted on 17 Jun 2015

Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands

Susanne Fuchs

Zentrum für Allgemeine Sprachwissenschaft, Berlin (Germany)

Caterina Petrone

Laboratoire Parole et Langage, CNRS UMR 7309 and Aix-Marseille University,

Aix-en Provence (France)

Amélie Rochet-Capellan

GIPSA-lab / Département Parole et Cognition, CNRS UMR 5216, Grenoble (France)

Uwe D. Reichel

Institut für Phonetik und Sprachverarbeitung, München (Germany)

Laura L. Koenig

Haskins Laboratories, New Haven & Long Island University, New York (USA)

Corresponding author

Susanne Fuchs

ZAS/Phonetik

Schützenstrasse 18

10117 Berlin

Email: fuchs@zas.gwz-berlin.de

Phone: ++49 30 20192 569

Fax: ++49 30 20192 402

Abstract (195 words)

Many past studies have sought to determine the factors that affect f0 declination, and the physiological underpinnings of the phenomenon. This study assessed the relation between respiration and f0 declination by means of simultaneous acoustic and respiratory recordings from read and spontaneous speech from speakers of German. Within the respective Intonational Phrase unit, we analysed the effect of the number of syllables and voiceless obstruents. Both factors could influence the slope of either f0 declination or rib cage movement. If respiration and f0 declination are related physiologically, their relationship might also be modulated by either one or both factors. Our results show consistently for both speech tasks that the slope of the rib cage movement is not related with f0 declination when length and consonant content vary. Furthermore f0 slopes are generally shallower in spontaneous than in read speech. Finally, although a higher number of voiceless obstruents yielded a greater rib cage compression, it did not affect f0 declination. These results suggest that although f0 declination occurs in many languages, it might not have a purely physiological origin in breathing, but rather reflects cognitive processing which allows speakers to look ahead when planning their utterances.

Acknowledgements

Introduction

This paper assesses the relationship between changes in f0 over time and a non-invasively obtained measure of time-varying respiratory system support for speech. Specifically, we evaluate the extent to which f0 declination varies with changes in rib cage volume. As discussed in more detail below, the degree to which f0 declination can be ascribed to respiratory vs. laryngeal mechanisms has been a subject of debate among past authors. We also explore whether any such relation is robust across utterance length and speaking tasks (reading versus spontaneous speech) that (a) vary in the demands placed on the respiratory system or (b) may lead to more or less frequent f0 declination.

On the universality and nature of declination

Broadly speaking, "declination" is the gradual decrease of f0 over the course of an utterance (e.g., Pike, 1945; Cohen &'t Hart, 1967; Maeda, 1976; 't Hart, Collier & Cohen, 1990), which has been observed especially in statements. Some aspects of declination may vary across typologically different languages and dialects (Yuan & Liberman, 2010; Schmid, Gendrot & Adda-Decker, 2012; Yoshida, 2011). Nevertheless, f0 declination has been reported for many tone and intonation languages. Among the intonation languages, f0 declination is found in Italian (Avesani, 1987), French (Collier, 1985; Schmid, Gendrot & Adda-Decker, 2012), English (Pierrehumbert, 1979; Lieberman, 1967; Maeda, 1976; Ladd, 1984), Dutch (Cohen & t'Hart, 1967; Strik, 1994; Strik & Boves 1995; Haan, 2002), Greek (Arvaniti, 2003), Danish (Thorsen, 1980), Swedish (Swerts, Strangert & Heldner, 1996), Spanish (Prieto, Shih & Nibert, 1996), Japanese (Yoshida, 2011), and German, the language under investigation in this paper (Oppenrieder, 1988; Grabe, 1998). Indeed, f0 declination may not be exclusively human, since it has been reported in calls of wild vervet monkeys and rhesus macaques (Hauser & Fowler, 1992).

The literature on declination phenomena has been characterized by varying terminology

and multiple theoretical perspectives (cf. Connell, 2001). Decreases in f0 may occur over a variety of time frames, leading to distinctions for some authors among global f0 changes such as *declination* and more local patterns such as *downstep* (usually used to refer to stepwise lowering of successive high tonal targets, e.g., Pierrehumbert, 1980; Liberman & Pierrehumbert, 1984) and *final lowering* (i.e., the extra f0 lowering of the utterance-final accent compared to its predicted value). Moreover, declination has been treated in phonetic as well as phonological terms (e.g., Ladd, 1984, 2008; 't Hart et al., 1990; Fujisaki, 1983, Thorsen, 1980, 1985).

In this paper, we follow Strik and Boves (1995) in using the term f0 declination to indicate a long-term decrease of f0 over large (multisyllabic, multiword) speech units. One reason for this focus on global changes is that thoracic volume and subglottal pressure signals generally show a steady decrease over long time frames rather than abrupt changes (cf. also Strik & Boves, 1995). We will evaluate the degree to which long-term f0 changes covary with changes in respiratory system support. As described in the next section, authors have disagreed on the strength of such a relationship.

Potential underlying mechanisms for declination

The observation of f0 declination across many languages and potentially even across species has led some authors to suppose that f0 declination is universal (e.g., Hombert, 1974; Ohala, 1978; Connell, 2001), and a natural outgrowth of speech system physiology. A frequent question has been whether f0 declination is primarily determined by respiratory factors (viz., subglottal pressure) or by activity of the intrinsic laryngeal muscles, both of which are known to affect vocal-fold vibrational frequency (e.g., van den Berg, 1958; Titze, 1989; Titze & Durham, 1987)[1] In an early study, Lieberman (1967) found a correlation between subglottal pressure and f0 declination in a sample of three subjects, and proposed that f0 declination

---

[1] Although less widely-studied, the extrinsic laryngeal muscles have also been shown to have possible effects on f0 (Maeda, 1976: 197ff.; Erickson, Baer & Harris, 1983).

might be a passive result of decreasing subglottal pressure over the course of an utterance. In his study, Lieberman assumed that the larynx muscles maintained fairly consistent levels of activity. Several subsequent studies explicitly investigated the role of laryngeal mechanisms in addition to subglottal pressure. Theoretical considerations also suggested that f0 declination might arise from a combination of factors. For one, the typical amount of subglottal pressure drop across an utterance may not fully explain the magnitude of observed changes in f0 (Titze, 1989). Ohala (1978) has further argued that laryngeal mechanisms (viz., changes in vocal-fold tension) are likely to play a role in f0 declination because declination serves to signal clause and sentence boundaries (e.g., Nakajima & Allen, 1993).

In order to tease apart the sources of f0 declination, Gelfer, Harris and Baer (1987) recorded acoustics, breathing kinematics, subglottal pressure and laryngeal muscle activation in two speakers of Dutch. One aspect of their experiment was designed to vary respiratory system demands and measure the potential consequences on f0 declination. Specifically, they used speech material with and without voiceless obstruents (reiterant speech using the syllables /ma/ vs. /fa/). Reiterant productions of /fa/ should lead to greater airflow expenditure than /ma/ and thus, to a greater loss of air over the course of an utterance and a steeper rip cage slope. Their results showed that the /fa/ utterances did indeed lead to a steeper rip cage slope in comparison to /ma/. However, f0 did not change accordingly and showed no effect due to the steeper rip cage slope. Thus, Gelfer et al. concluded that subglottal pressure decrease is actively controlled (rather than passively decreasing), and f0 patterns can compensate for the loss of air.

Strik and Boves (1995) recorded laryngeal activity, subglottal pressure and acoustics in two speakers of Dutch. Although they did not exclude the possibility that f0 declination might be caused by laryngeal muscle activity, they pointed out that long-term downward trends were evident in f0 and subglottal pressure, but such global patterns did not seem to characterize laryngeal muscle activity. They also provided counterarguments to both of the considerations

above regarding laryngeal control of declination: They noted that measures of f0 decrease over an utterance varied highly across methodologies, and reviewed the considerable evidence for respiratory system control during speech; that is, control was not limited to the larynx. As such, they concluded that one source of f0 declination is subglottal pressure.

A different account of f0 declination was proposed by Maeda (1976). In x-ray recordings of one subject, Maeda found a shortening of the vocal folds towards the ends of utterances and a lowering of f0. In a follow-up study, Maeda simultaneously recorded acoustics and electromyographic data from two subjects' cricothyroid (CT), vocalis (VOC), and lateral cricoarytenoid (LCA) muscles. Results showed that the previously observed shortening of vocal folds might not result from the activation of these intrinsic laryngeal muscles. Consequently Maeda proposed the *tracheal pull hypothesis*: Speech production relies on expiratory airflow, which involves a decrease in thoracic volume (rib cage compression) over the course of an utterance. This decrease lowers the position of the sternum, thereby increasing tracheal pull on the cricoid cartilage, leading to vocal fold shortening and thus, lowering of f0 (see Watson, Ciccia & Weismer, 2003 for additional evidence for the tracheal pull hypothesis).

Physiological data elucidating the mechanisms of declination are rather sparse, partly owing to the invasiveness of direct methodologies for evaluating muscular control and respiratory system support. Direct assessments of laryngeal muscle activity can only be done using electromyography, which typically involves anesthetization and carries risks of nerve damage. Direct assessment of subglottal pressure mainly relies on tracheal puncture (e.g., Lecuit & Demolin, 1998); other authors have had speakers swallow a balloon into the oesophagus to measure tracheal pressure indirectly (Draper, Ladefoged & Whitteridge, 1960; Slifka 2003). Less invasive methods include estimating subglottal pressure from intraoral pressure plateaus (Löfqvist, Carlborg & Kitzing, 1982), and sudden airflow interruptions by means of a mouthpiece (Hoffman & Jiang, 2009; Hoffman, Baggott, & Jiang, 2009); the

drawback of these is that they severely limit the speech material. These constraints motivated us to investigate the potential impact of the respiratory system on f0 declination by means of Inductance Plethysmography (formerly Respitrace). This non-invasive methodology allows us to study the physiological mechanism of respiration during speech production in combination with the acoustics in a relatively large number of subjects. The participants of the study produced rather large corpora in relatively natural settings which increases the ecological validity of the results.

Sources of variation in f0 declination

Although f0 declination has been widely observed across languages and perhaps even across species, it is clearly not mandatory for all utterances. For instance, f0 declination may not occur across all sentence types. Declination has been studied mainly in simple declarative sentences and is thought to signal finality, with the f0 contour usually represented as a negative slope (i.e., declining regression line) over the whole sentence. F0 declination has been claimed to apply in varying degrees to other sentence types. For instance, Thorsen (1980) suggested for Danish, and van Heuven & Haan (2002) for Dutch, that declination is very steep for declarative utterances, less steep for non-terminal statements and it is suspended or even positive (i.e., rising regression line) in echo questions and in utterances that are morphosyntactically statements but are intonationally marked as questions.

Of importance to the current work, studies have also demonstrated that the degree of f0 declination is influenced by three other factors: the speech task, the length of the utterance, and the language.

Lieberman et al. (1985) investigated the effect of the speech task. They first selected 19 unemphatic declarative sentences shorter than 3.5 seconds, taken from recordings of spontaneous speech. These sentences were subsequently read by seven American English speakers without any specific instructions. The results showed a greater variability in the f0

declination slope in the spontaneous speech data than in the read speech, and more declination, i.e. a higher percentage of negative slopes, in read than in spontaneous speech.

Anderson and Cooper (1986) carried out a comparable experiment where five American English speaking subjects spontaneously produced five sentences with one or two clauses in picture naming tasks. Afterwards, the speakers also read these five isolated sentences aloud. In contrast to Lieberman et al. (1985), Anderson and Cooper found no significant difference between the slopes of f0 declination in the two speech tasks. However, utterance length did show a significant effect; i.e. steeper f0 declination was found for shorter sentences in comparison to longer ones. Anderson and Cooper (1986) concluded that declination is independent of the speech task, but they also noted the possibility that f0 declination occurs more frequently in isolated sentences (like the ones they recorded), than in longer passages.

The effect of utterance length on f0 declination has been replicated by several authors in different languages (e.g., Cooper & Sorensen, 1981 and Ohala, 1978 for English; Pierrehumbert & Beckman, 1988 for Japanese; Swerts, Strangert & Heldner, 1996 for Swedish; Grabe, 1998 for German and English). Authors have noted that declination may be an acoustic parameter reflecting anticipatory speech planning (e.g., Kochanski & Shih, 2001; Ohala, Dunn & Sprouse, 2004; Horne, Frid & Roll, 2006; Yuan & Liberman, 2010). Thus, it could provide a window into this complex cognitive process.

Language effects on f0 declination have been reported by various authors. Schmid, Gendrot and Adda-Decker (2012) compared French and German in large broadcast speech corpora. To compare the two languages in a reliable manner, the authors excluded a considerable amount of data: They chose utterances between one and four seconds. Shorter utterances were excluded, since f0 declination is highly influenced by local accentuations if the utterances consist only of a few words. Moreover, only decreasing f0 contours, i.e. negative f0 slopes, were taken into account, since f0 declination is commonly defined precisely as the downtrend in f0 (e.g.,Strik & Boves, 1995). These exclusion criteria resulted

in a final corpus of 36.4 % of the original data for French speakers and 40.2% for German speakers. This result implies already the more limited use of f0 declination in naturalistic, spontaneous speech than in read speech recorded in the laboratory. Although the French and German data generally showed a decrease of 2.5 semitones per second, significant cross-linguistic differences were reported in the f0 top- and bottom-lines, suggesting that these may be more sensitive to language specific constraints due to prosody, and accentuation.

Yuan and Liberman (2010) also compared broadcast speech in a tonal (Mandarin Chinese) and a non-tonal language (English). Besides the prominent effect of utterance length on f0 declination described above, the authors show that Chinese and English differ particularly in top- and baseline properties. Whereas the midline curves corresponded to a linear regression fit to all f0 values, top- and baseline values were defined as local peaks and valleys based on a convex hull curve. The toplines only included the peaks of the respective utterance, the baselines only the valleys in a given utterance. The results showed that top- and baselines had rather similar slopes in English, but very different in Chinese (flat for the bottom-line and steep for the topline). The authors therefore concluded that top- and baseline measures are sensitive to language specific properties and may not be an automatic consequence of a physiological process.

Aims and hypotheses

The general aim of this study was to investigate the relationship between f0 declination and changes in rib cage volume, and the ways in which these measures and their interrelationships might be affected by utterance characteristics. The tracheal pull mechanism (Maeda, 1976; Watson, Ciccia & Weismer, 2003) would predict a relation between the two measures. If f0 declination is a consequence of the respiratory system, we would expect that rib cage volume covaries with f0 slope. We would also expect that f0 slopes change when respiratory dynamics vary as a function of changing demands on the system. As a test bed for these

hypotheses we used two factors, the number of syllables and the number of voiceless obstruents in an Intonation Phrase. Several studies (e.g. Cooper & Sorensen, 1981; Ohala, 1978; Kochanski & Shih, 2001; Yuan & Liberman, 2010) reported that the f0 slope varies as a function of the number of syllables. They found flatter f0 slopes for a greater number of syllables in an utterance. If f0 declination and respiration are linked, rib cage slopes should also be flatter in longer utterances. As a second factor, we assessed whether rib cage slopes and f0 declination varied with the number of voiceless obstruents, which are associated with high airflow rates (e.g., Klatt, Stevens & Mead, 1968; Stathopoulos & Weismer, 1985). Gelfer, Harris and Baer (1987) found an effect of voiceless obstruents on rib cage slopes, but did not find that the number of voiceless obstruents affected f0 declination measures. However, their number of speakers was small and their results were based on reiterant speech, which is rather unnatural. Since the vocal-fold abduction associated with voiceless obstruents increases airflow expenditure, we hypothesized that utterances with many voiceless obstruents might lead to a faster reduction of rib cage volume and a steeper f0 declination.

Lastly, we aim at a better understanding of the extent to which f0 declination varies with speech task, following on Lieberman et al.'s (1985) finding of steeper f0 declination for read than spontaneous speech. For this reason we recorded read and spontaneous speech. We also extended previous work by not only recording single sentences, but longer passages of speech.

Using non-invasive Inductance Plethysmography allowed us to evaluate these questions in a fairly large number of speakers and utterances, and helped ensure speech naturalness.

Methodology

General experimental set-up and procedures

Participants were graduate or undergraduate students who reported no history of speech, language or hearing difficulties and no heavy smoking habits (i.e., more than ten cigarettes per day). All spoke Standard German. Acoustic and breathing data were simultaneously

recorded while subjects were standing. Paper sheets with the text to be spoken were put on a music stand (see Figure 1A) and pages were turned by one of the experimenters, since the subject was instructed not to move her arms or legs during speech to avoid distortions of the breathing signals. All subjects were selected based on a normal Body-Mass-Index (BMI) ranging from 19-25, given that excess body fat can disturb the respiratory signals (Parameswaran, Todd & Soth, 2006).
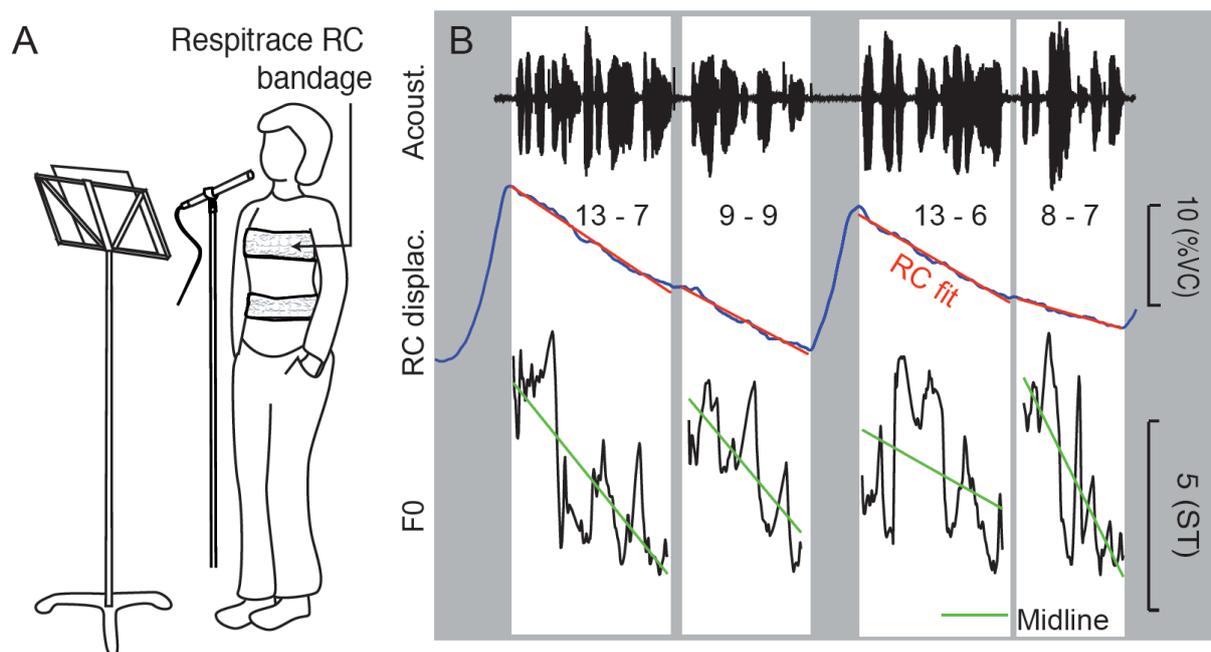

Figure 1 A. Experimental set-up B. Example of signals (acoustics - top row, rib cage displacement - middle row and f0 - bottom row). (B) Examples for the reading the texts: "Diese Güte wurde später von der Maus belohnt" (Translation: *This kindness was later on rewarded by the mouse)* [13 syllables, 7 voiceless obstruents] "so unwahrscheinlich es zunächst klingt" (*so unlikely as it may sound*) [9 syllables, 9 obstruents], "Eines Tages fing sich der Löwe in einem Netz (*One day the lion was caught up in a net*) [13 syllables, 6 obstruents], "das als Falle aufgestellt war" (*which was positioned as a trap)* [8 syllables, 7 obstruents]. Superimposed lines on the RC displacement and f0 represent the linear fits used to evaluate rib cage exhalation slope and f0 declination respectively.

Respiratory movements were monitored by means of Inductance Plethysmography. For this technique the participants wear two elasticized bands 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), as represented on Figure 1A. The bands enlarge their circumference during inhalation and reduce their circumference during exhalation, where speech is produced. Expansion and compression of the bands is measured via the electrical resistance of small wires located in the two bands. For this study, only rib cage movements were considered. In more than half of the subjects we could not reliably label abdominal movements because their amplitude was too small. Changes in rib cage movement are often described as changes in rib cage volume and we will follow the use of this terminology. We note that this is a simplification, since the Respitrace signal provides only two-dimensional information. All data were recorded with a sampling frequency of 11030 Hz.

At the end of the recording for every speaker, vital capacity (VC) manoeuvres were performed in order to evaluate subjects' maximal rib cage compression and expansion. This measure was necessary to express rib cage movement in percentage of VC for each subject (e.g., Huber, 2007) and to allow between-subjects comparisons of rib cage changes.

Speakers and stimuli

Ten female native speakers of German, ages 21-30 years, were recorded. The data set was a subset of a larger project presented in Rochet-Capellan and Fuchs (2013). The read speech material consisted of 15 short texts (fables from Lessing, Aesop, and Fontane), including 80 to 132 words (mean 95.4). Subjects were first instructed to read five of these texts. In a following step they were instructed to listen attentively to acoustic recordings of the ten remaining texts read by a female or a male speaker and to summarize the story afterwards

using their own words. The semantic content of the texts is therefore comparable between the read and the spontaneous speech data.

Labelling of the acoustic data

The Intonational Phrase (IP) was chosen for two reasons. First, it has been suggested that the domain of f0 declination in German is the IP (Wiese, 2000). Second, we expect that if any relation between breathing and f0 declination exists, it would be evident in large prosodic domains, since speech breathing cycles are relatively long in duration. In a previous study (Rochet-Capellan & Fuchs, 2013b) we found that breathing cycles during spontaneous speech production of 26 female subjects lasted on average 3.5 seconds and consisted on average of 16 syllables.

The annotation of IP boundaries was separately performed by two authors (SF and CP), who are both trained in prosodic labelling of acoustic corpora. The annotation was based on both perceptual and acoustic criteria: perception of strong prosodic boundaries and presence of boundary tones, final lengthening and silent pauses. We are aware that a more detailed intonational analysis could contribute to investigating possible interdependence between global f0 trends (i.e., declination) and local phonological events (e.g. downstep across pitch accents). However, subjective coding of perceptual prominence is partly influenced by linguistic knowledge and it may show lower inter-rater agreement than, for instance, the perceptual coding of strong boundaries (Cole, Mo & Hasegawa-Johnson, 2010; Baumann, 2014; Petrone & Truckenbrodt, in press). Finally, it is still problematic to establish how perceptual prominence relates to specific acoustic events and prominence parameters, especially in spontaneous speech (cf. Wagner, Tamburini & Windmann, 2012 and references therein). As a consequence, we decided to focus only on IP labelling.

The labels for each token were stored in Praat Textgrid files (Boersma & Weenink, 2011). Labellers' reliability was first evaluated by measuring the inter-transcriber agreement

(Pitrelli et al., 1994). Specifically, the labellers began by separately annotating a randomly selected subset of the data (21.3% of all data, consisting of two texts from the read and two summaries from the spontaneous speech dataset for eight out of ten subjects). For the read speech corpus the agreement score was 90.8% and for the spontaneous speech corpus it was 98.4%. Given this high level of agreement, the two annotators each proceeded to label half of the remaining read and spontaneous speech corpora.

Pre-processing of f0 and its declination

For all data f0 was measured using Praat with a time step of 0.010 seconds and a frequency range between 100 and 500 Hz. Figure 1 (B & C, bottom row) displays an example. The linear fit (midline slope) used to analyse f0 declination is plotted as a straight line over the course of the IP as defined by the acoustic/perceptual labelling.

The measurement of f0 declination was obtained as follows: For speaker normalization an f0 base value *b* was calculated as the median below the 5th percentile. F0 was then transformed to semitones relative to this base value as:

$$f0 \text{ [st]} = 12 * \log\_2 (f0 \text{ [Hz]} / b).$$

After this pre-processing a declination midline was fitted to the f0 data by linear regression in each IP. Midlines were chosen in this study for several reasons. They have been extracted as correlates of f0 declination in several studies (e.g., Lieberman et al., 1985 and Swerts et al., 1996). Midlines are more robust than top- and baselines (Yuan & Liberman, 2010). Perceptual studies have found that equivalent prominences for f0 movements varying in magnitude are not judged relative to the baseline, but rather a reference line located more centrally within the f0 range (Liberman & Pierrehumbert, 1984; Ladd, 1993). Two more considerations support the choice of looking at the global f0 slope instead at the relation between local peaks and valleys. First, the acoustic detection of peaks and valleys can be a very difficult task, given that (1) microprosodic perturbations can mask the targets and (2)

phonologically relevant high and low targets are not always instantiated as f0 turning points (e.g., Pierrehumbert, 1980). Moreover, we prefer analyses of articulation and prosodic structure that are based on the whole time series instead of selecting single time points which are highly based on theoretical assumptions (for a brief discussion of "magic time point analysis", see Mücke, Grice & Cho, 2014).

Processing of breathing data

Rib cage signals were pass band filtered over 1-40 Hz with a 50 dB damping factor using a finite impulse-response (FIR) filter in order to remove traces of electrical interference. As can be seen in Figure 1, the respitrace signals primarily show long-term or global variation; local events are quite small in magnitude. The signals were then resampled at 200 Hz to reduce the size of storage while preserving the local events that could occur during the exhalation phase. Finally each speaker's signals were normalized relative to her vital capacity (VC), which allows expressing changes in amplitude relative to speakers' respective percentage of VC as well as the rib cage slope in %VC/s. The slope of rib cage compression during speech production was computed using a linear fit over the rib cage data during the acoustically defined time intervals (cf. the superimposed line on the breathing signal on figure 1B and 1C, middle row).

Syllable, voiceless obstruent and number of clauses counts

Syllable and voiceless obstruent counts were derived automatically from the output of the BALLOON toolkit (Reichel, 2012), which provides a syllabified canonical German grapheme-to-phoneme conversion. The conversion model is implemented as a C4.5 decision tree (Quinlan, 1993) trained on grapheme, morpheme, and part of speech features. The syllabification of the transcription is done in two steps. Syllable boundaries are placed in front of each sonority minimum, and their locations are subsequently adjusted for violations of

German phonotactics as specified by Kohler (1995). In the resulting syllabified canonical transcription we counted syllables and phonologically voiceless obstruents.

Variables and statistical analyses

The statistical significance of the results was tested using a series of linear mixed models by means of the R-environment (R Development Core Team, 2012). The mixed models were run on the f0 slope and the rib cage slope, separately for the read and spontaneous speech data. We included both negative and positive f0 slopes in the statistical analysis. Positive f0 slopes were considered as a control condition for which we would not expect a relation between f0 slope and rib cage slope.

The models included number of syllables (Syll), number of voiceless obstruents (Obstr) and direction of the f0 slope (negative vs. positive) as fixed factors. Random slopes were included for the fixed factors (Barr et al., 2013). The factor Syll was centered at its minimum value (Baayen, 2008). Centering is recommended because otherwise the model would set the intercept at 0 as a default value. However, there are no utterances with zero syllables. Possible relationships between the f0 midline and rib cage slope were tested by introducing each of them as covariates in the models. Because of the complexity of the multiple analyses performed on the datasets, only pairwise interactions were considered.

Speakers and Text were the random intercepts, since we were not interested in effects of speaker- and text-specific differences in the production patterns. The random coefficients were uncorrelated in order for the model to get a higher number of degrees of freedom.

Since the full models showed some over-parameterization, backward elimination based on graphical inspections and on likelihood-ratio tests was used to decide which terms should be retained in the models. Likelihood-ratio tests (Baayen, 2008) were run comparing nested models, that is more complex models (e.g., a model which contained a specific random slope) with simpler ones (e.g., without that random slope).

A moderate correlation of the factors Syll and Obstr was found, since increasing the number of syllables leads to an increase of number of obstruents ($r^2 = 0.52$ for read speech; $r^2 = 0.59$ for spontaneous speech, cf. Figures 2). To better isolate the contribution of each factor, full models containing both factors were compared with simpler models containing either one or the other factor. Such comparisons were performed by an additional series of likelihood ratio tests. If the simpler model (e.g., containing only the factor Obstr) was significantly different from the full model (containing both Syll and Obstr), we concluded that both factors are required in the model. If the simpler model was not significantly different from the full one, we concluded that the additional factor was not required. The effect of adding or dropping terms was also explored using the Akaike's Information Criterion (AIC) value of the models (cf. Venables & Ripley, 1999). The AIC provides a weighted measure of the quality of a statistical model, by adjusting its goodness of fit to the number of factors included in the model. The model which has the lowest absolute AIC value is considered to be the most parsimonious one and is then selected in line with Ockham's razor.

Bootstrap methods with replacement from the original sample (number of samples = 1000) were employed to estimate 99% confidence intervals (CI) for each mixed model (Efron & Tibshirani, 1986). For likelihood ratio tests comparing mixed models, the cut-off point for significance was set at $p < 0.01$. In what follows, results of the statistical analyses on the two datasets are reported in connection with our experimental questions.
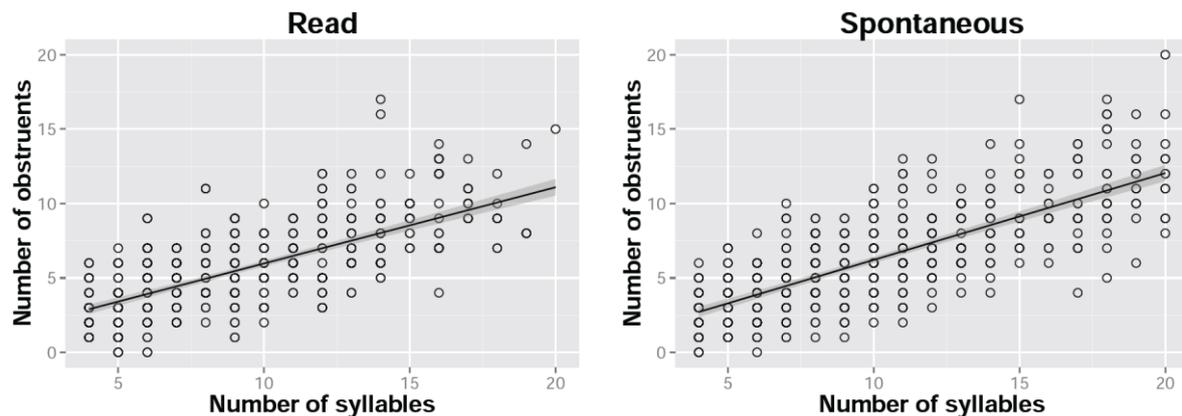
Figure 2. Correlation between number of voiceless obstruents and number of syllables within IPs for the read speech (left) and spontaneous speech data (right). The straight lines indicate the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.

Data selection and data set

Table 1 summarizes the excluded data. Exclusion criteria were the following:

1) Number of syllables < 4 and > 20. Based on graphical inspection of the histogram and the density curve for the number of syllables we excluded very short and very long IPs. This allowed us to obtain a more symmetric distribution of the tokens across the number of syllables. Various studies investigated f0 declination for utterances longer than 750 ms − 1 second (e.g., Nakajima & Allen, 1993; Yuan & Liberman, 2010), since the shorter the utterance, the higher the impact of local f0 variations on the f0 slope.

2) Null f0 values in more than 50% of the samples. We excluded from the analysis IPs in which f0 values were available in fewer than 50% of samples (e.g., due to the presence of a high number of voiceless consonants). In these cases, f0 contours could not be reliably obtained through f0 interpolation.

Table 1: Summary of the excluded data.

Exclusion criteria	Read speech (out of 834 tokens)	Spontaneous speech (out of 1033 tokens)
Number of syll < 4	81	276
Number of syll > 20	25	45
F0 $< 50\%$	25	56
Analysed data (in % of all)	703 (84.3%)	656 (63.5%)

Results

Is there a relation between f0 declination and the respiratory system?

In Figure 3 the relation between the f0 slopes and the rib cage slopes (in per cent of vital capacity) is displayed for the two datasets. The data are split by the f0 slope direction (negative vs. positive). The figures suggest that there is no relationship in any of the two datasets. The statistical analysis confirmed that the relation between the two parameters was not significant. In general the slope of rib cage movement is much more dispersed than the slope of f0 declination. Furthermore, the data from the different speech tasks seem to overlap to a large extent.
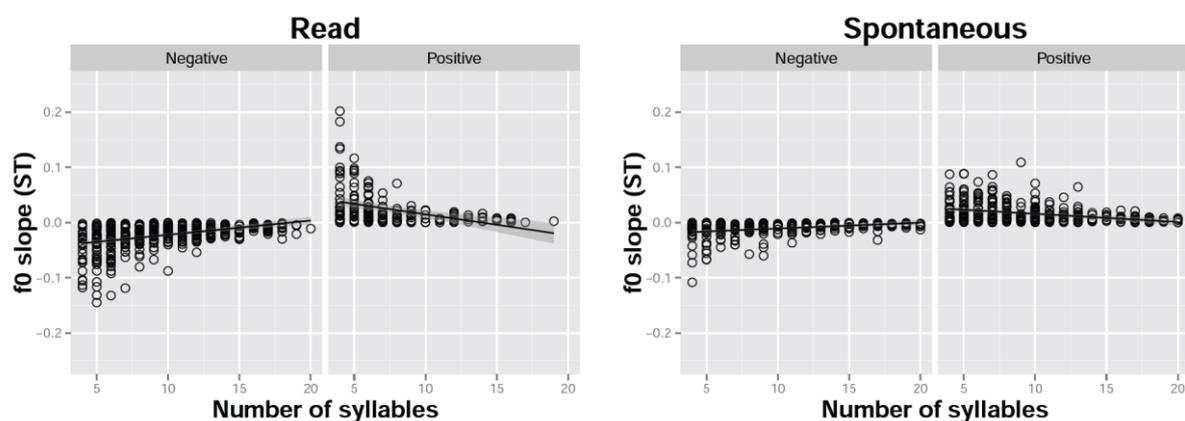
Figure 3. Relation between f0 midline slopes in semitones (y-axis) and rib cage slopes (x-axis) for the read (left) and spontaneous (right) speech data. In each panel, data are split by the direction of the f0 midline slope.

Does f0 declination and rib cage slope depend on the number of syllables to be produced?

Figure 4 shows the f0 slope values plotted against the number of syllables (from four to twenty) in the IPs, separated for the two datasets. Both figures suggest a relationship between f0 slope and number of syllables, in that f0 slopes are on average steeper in IPs with fewer syllables than in those with many syllables, independently of the direction of the f0 slope. For read speech, the mean negative slope is -0.054 at IPs containing four syllables and 0.011 (i.e., it is f0 declination contour is flat) at IPs with 20 syllables. The mean positive f0 slope is 0.053 at IPs with four syllables and 0.002 at IPs with 20 syllables. The difference in slope across number of syllables is smaller for spontaneous speech. For this dataset, the mean negative slope is -0.030 at IPs with four syllables and -0.005 at IPs with twenty syllables; the mean positive slope is 0.034 at IPs with four syllables and 0.003 at IPs with 20 syllables. This means that f0 slopes in short IPs are shallower in spontaneous than in read speech.

Moreover, the variability of the f0 slope is greater for IPs containing few syllables. For instance, in the read speech, the standard deviation is around 0.054 at IPs (averaged over positive and negative slopes) containing four syllables but it is close to 0 at IPs with 20 syllables in both datasets. The statistical analysis confirmed that IPs with many syllables had smaller f0 slope coefficients, i.e., a flatter f0 slopes. This effect was significant only for negative slopes, for both read [β = 3.462e-03, SE = 8.894e-04, t = 3.892, p = 0.001, CI = 9.739e-04; 5.7875e-03] and spontaneous speech [β = 1.273e-03, SE = 5.801e-04, t = 2.194, p = 0.03, CI = -3.382e-04; 0.002]. Neither the interaction between number of syllables and number of voiceless obstruents nor between number of syllables and rib cage slope was

significant. For the rib cage slope, no significant effects of the number of syllables were found in either dataset. Finally, a direct comparison was made between read and spontaneous speech to statistically verify the difference in the f0 declination slope across speech styles. This additional model was run only on negative slopes, in which the effect of the number of syllables was significant. The model confirmed that f0 declination is globally shallower in spontaneous than in read speech [β = 2.322e-02, SE = 2.322e-02, t = 4.8, p = 0.001, CI = 0.0094329; 3.584e-02].
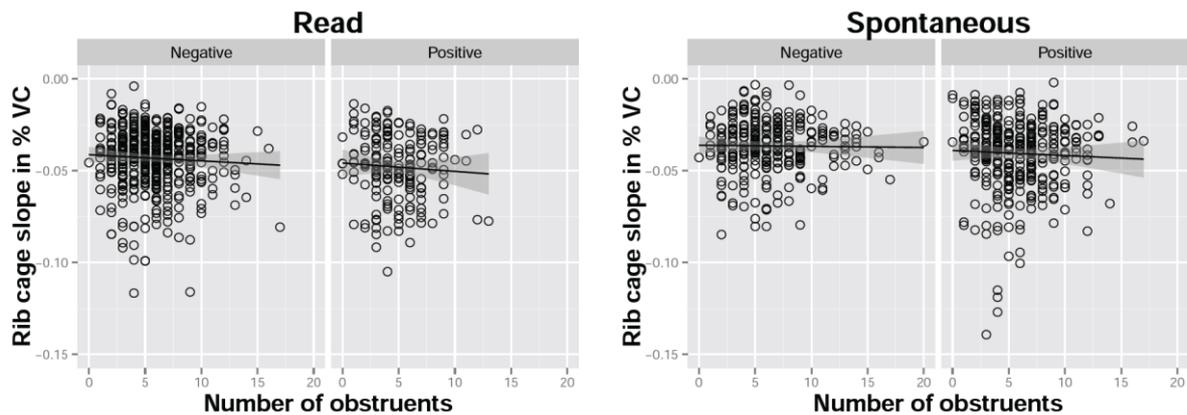

Figures 4. Relation between f0 midline slope and number of syllables within the IPs for the read (left) and spontaneous (right) datasets. The straight line indicates the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.

Does the number of voiceless obstruents influence rib cage slope and f0 declination?

Figure 5 shows the rib cage slope against the number of obstruents (Obstr). Despite the large variability, a greater number of voiceless obstruents led to slightly steeper slopes especially in the read speech data set. For instance, the rib cage slope was -0.044 when the utterance contained only one voiceless obstruent and it was -0.080 for an utterance containing 17 voiceless obstruents. The effect of Obstr in the read speech was significant, but only for

negative slopes [β = -2.103e-03, SE = 5.531e-04, t = -3.8, p = 0.001, CI = -3.363e-03; -6.703e-04], as we expected. No effects were found in the spontaneous speech data set.

Crucially, the number of obstruents did not influence the f0 slope, even in the case where it affected the rib cage slope.

Figures 5. Relation between rib cage slope and number of obstruents within the IPs for the read (left) and spontaneous (right) datasets. The straight line indicates the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.

What is the independent contribution of the number of voiceless obstruents and number of syllables?

We ran likelihood ratio tests to evaluate the impact of the number of syllables and voiceless obstruents on rib cage and f0 midline slope since the two factors were moderately correlated. For rib cage slopes, the likelihood ratio test showed that, for read speech, there was no significant difference between the full model containing both Syll and Obstr and the simple model only containing the factor Obstr. However, fitting a full model was significantly better than fitting a model containing only Syll (p < 0, AIC= -3697.3). This confirmed that Obstr (and not Syll) is a good predictor for rib cage slope. For the spontaneous speech data,

however, there was no significant difference between the full and the simpler models. This is in line with the fact that neither Syll nor Obstr affected rib cage slope in spontaneous speech.

For f0 midline slope, there was no difference between the full model and a model containing only Syll. However, fitting a full model containing both Syll and Obstr was significantly better than the fitting of a model containing only Obstr, both for read ($p < 0$, AIC= -3533.4) and spontaneous speech ($p < 0$, AIC = -3697) data. This means that Syll (and not Obstr) is a good predictor for f0 declination. We conclude that the effect of Syll on f0 declination is genuine.

Discussion

Our study provides evidence that f0 declination behaves independently of changes in rib cage compression. This evidence derives from the following findings:

First, whereas f0 slope varied as a function of the number of syllables, the rib cage compression did not change accordingly. Although it has been reported previously (Whalen & Kinsella-Shaw, 1997; Fuchs et al., 2013; Rochet-Capellan & Fuchs, 2013b) that respiration changes with respect to the number of syllables in an upcoming utterance, this is only reflected in inhalation depth and duration, not in the exhalation slope.

Second, although rib cage slope is affected by the number of voiceless obstruents produced in an Intonation Phrase (as we found for read speech data) it has no consequence on the f0 slope. The lack of effect of the number of voiceless obstruents on f0 slope also confirms the work by Gelfer and colleagues (Gelfer, Harris & Baer, 1987), and extends it to a larger dataset, a greater variety of voiceless obstruents, and more natural speech tasks. It suggests that f0 declination is independent of respiration and that even if voiceless obstruents may lead to a steeper rib cage compression, as was found in our read speech dataset, speakers can (and do) compensate for more extreme losses of air so that f0 declination stays unaffected.

Third, there is no relation between f0 declination and rib cage compression across different speech tasks, even though slope values differ between the two tasks (read speech showed steeper f0 downward slopes than spontaneous speech data).

It should be noted that the respiratory data we obtained cannot be considered to be a mirror of subglottal pressure; rather, continuously decreasing lung volumes in speech are typically associated with subglottal pressures that can remain fairly constant (e.g., Slifka, 2003). That is, we may not discount a role for subglottal pressure in f0 declination. Our results do, however, make the tracheal pull hypothesis proposed by Maeda (1976) rather unlikely, since we found a large scatter of rib cage changes while f0 slope was relatively stable. It appears that a laryngeal explanation for declination might not be so plausible either given that laryngeal muscle activation generally does not span long temporal intervals (e.g., Strik & Boves, 1995). Thus, although declination has been widely observed across languages, the underlying physiological mechanism remains a question.

The effects of speech task seen here suggest that f0 declination reflects anticipatory speech planning or prepared speech. Most of the studies we are aware of used read speech material where subjects can anticipate the upcoming text. Hence, f0 declination may not be universal across communicative conditions, but may be limited to certain speech tasks (viz., reading). We found that if f0 declination occurs, then it can signal the length of the upcoming utterance. That is, German speakers know in advance how long the IP will be and adjust the declination rate on the length of the prosodic domain: When an IP is relatively long, the slope of f0 declination is flat whereas it is steeper when the IP is short. This result is consistent with the literature (Anderson & Cooper, 1986; Cooper & Sorensen, 1981; Ohala, 1978; Swerts, Strangert & Heldner, 1996; Grabe, 1998). Speech tasks with more possibilities to look ahead (reading) show more frequent negative f0 slopes and steeper ones. Comparably, rising f0 slopes are also steeper for IPs with fewer syllables than for longer utterances.

We note though that linking f0 declination to speech planning is not a trivial task. Some previous studies have in fact claimed that look-ahead mechanisms are "psycholinguistically implausible" since planning long stretches of speech would require cognitive overload (cf. Pierrehumbert & Lieberman, 1984; Ladd, 2008).

However, recent empirical studies also suggested that the unit of planning is flexible, both depending on the task (prepared vs. unprepared speech; cf. Swets et al., 2013) and the speaker (e.g., differences in cognitive skills; cf. Wagner et al., 2010).

In our study, we also found that the effect of utterance length is modulated by speech task. This might indicate that planning of f0 declination depends on whether the speech is prepared or not. Self-paced reading tasks usually implies more "prepared" speech, since speakers might delay utterance production for a brief period of time in order to plan in advance, e.g., the prosody of the utterance. On the other hand, spontaneous speech is less prepared, so that speakers cannot have fully planned the prosodic structure in advance. In connection with that, it is noteworthy to note that previous studies have almost exclusively focused on read speech, and have found evidence for effects of utterance length on f0 declination (cf. Yuan & Liberman, 2002 for different results for f0 slope across speech tasks). We like to mention that our spontaneous speech data are monologues and subjects do not expect to be interrupted or that another interlocutor takes the turn. Conversational interactions may show more f0 declination, since a steep declination together with a final f0 lowering may signal the opportunity to take the turn (Beattie, Cutler & Pearson, 1982) or a steep declination may be used to signal the introduction of a new topic (Nakajima & Allen, 1993).

Moreover, indirect evidence of different planning strategies in our spontaneous speech dataset comes from the fact that it contained many IPs consisting only of one or two words. The high occurrence of short IPs might suggest that in spontaneous speech subjects may not always need to plan ahead as much as in read speech.

Summary and conclusion

In this paper, we investigated the potential relationship between respiratory system and f0 declination by means of simultaneous respiratory and acoustic recordings. Respiratory activity was monitored by means of Inductance Plethysmography, which allowed generalization across many speakers and different tasks.

Our results suggest that f0 declination is not determined by respiratory factors since its production is independent of both rib cage movements and changes in respiratory demands resulting from variation in number of obstruents. This also leads us to reject previous physiological accounts such as the Maeda's (1976) tracheal pull hypothesis, according to which a direct link is posited between rib cage compression and f0 lowering. The current findings lead us to conclude that respiratory contributions to f0 are rather limited; laryngeal control is an obvious alternative for future work to explore.

Further, we think that our results point to the existence of look-ahead mechanisms underlying the production of f0 declination. Specifically, f0 declination is modulated by the length of the prosodic domain in which it applies (the Intonation Phrase), the slope being steeper in read than in spontaneous speech. This is in line with recent psycholinguistic research claiming that incremental planning is under strategic control. Further research is needed to explore more fully possible physiological constraints other than thoracic volume changes on the production f0 declination as well as their interrelation with cognitively grounded differences among tasks.

References

Anderson, S.W., & Cooper, W.E. (1986). Fundamental frequency patterns during spontaneous picture descriptions. *Journal of the Acoustical Society of America*, *79(4),* 1172–1174.

Arvaniti, A. (2003). Peak scaling in Greek and the role of f0 declination. *Proceedings of the XVth ICPhS Barcelona*, 2269–2272.

Avesani, C. (1987). Declination and sentence intonation in Italian. In P.M. Bertinetto (ed.) *Quaderni del Laboratorio die Linguistica della Scuola Normale Superiore di Pisa, 1,* 8–23.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Baumann, S. (2014). The importance of tonal cues for untrained listeners in judging prominence. *Proceedings of the 10th ISSP*, Cologne (Germany), 21–24.

Beattie, G.W., Cutler, A. & Pearson, M. (1982). Why is Mrs. Thatcher interrupted so often? *Nature 300*, 744 – 747.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.3, retrieved from http://www.praat.org/

Cohen, A. R., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua, 19,* 177–192.

Cole, J., Mo, Y. & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Journal of Laboratory Phonology, 1*, 425‑452.

Collier, R. (1985). F0 declination: The setting and resetting of the baseline. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, 19,* 111–132.

Connell, B. (2001). Downdrift, downstep and declination. In U. Gut & D. Gibbon (eds). *Typology of African prosodic systems* (pp. 3–12). Bielefeld: University of Bielefeld.

Cooper, W., & Sorensen, J. (1981). *Fundamental Frequency in Sentence Production*. New York: Springer Verlag.

Draper, M.H., Ladefoged, P., & Whitteridge, D. (1960). Expiratory pressures and air flow during speech. *British Medical Journal*, *1,* 1837–1843.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1*: 54–77.

Erickson, D., Baer, T., & Harris, K.S. (1983). The role of the strap muscles in pitch lowering. In D.M. Bless & J.H. Abbs (eds.) *Vocal fold physiology: Contemporary research and clinical issue* (pp. 279–285). San Diego, CA: College-Hill Press.

Fuchs, S., Petrone, C., Krivokapic, J. & Hoole, P. (2013) Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41*, 29–47.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P.F. MacNeilage (ed.), *The Production of Speech* (pp. 39–55). New York: Springer.

Gelfer, C.E., Harris, K.S., & Baer, T. (1987). Controlled variables in sentence intonation. In T. Baer, C. Sasaki, and K. S. Harris (eds.) *Laryngeal function in phonation and respiration* (pp. 422–435)*. Boston, MA: Little Brown.

Grabe, E. (1998). *Comparative intonational phonology: English and German*. Doctoral dissertation, Radboud University Nijmegen.

Haan, J. (2002). *Speaking of questions: An exploration of Dutch question intonation*. Utrecht: LOT.

Hauser, M. D., & Fowler, C. A. (1992). Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustical Society of America*, *91*(1), 363–369.

Hoffman, M.R., Baggott, C.D., & Jiang, J.J. (2009). Reliable time to estimate subglottal pressure. *Journal of Voice, 23(2),* 169–174.

Hoffman, M.R., & Jiang, J.J. (2009). Estimating subglottal pressure via airflow interruption with auditory masking. *Journal of Voice, 23(6),* 653–657.

Hombert, J.-M. (1974). Universals of downdrift: Their phonetic basis and significance for a theory of tone. In W.R. Leben (ed.) *Papers from the Fifth Conference on African Linguistics, Studies in African Linguistics, Supplement* 5 (pp. 169–183).

Huber, J.E. (2007). Effect of cues to increase sound pressure level on respiratory kinematic patterns during connected speech. *Journal of Speech, Language, and Hearing Research, 50,* 621–634.

Klatt, D.H., Stevens, K., & Mead, J. (1968). Studies of articulatory activity and airflow during speech. In A. Bouhuys (ed.) *Sound Production in Man* (pp. 42–55). New York: New York Academy of Sciences, Vol. 55.

Kochanski, G. & Shih, C. (2001). Automated modelling of Chinese intonation in continuous speech. *Proceedings of Eurospeech, Aalborg (Denmark),* 911–914.

Kohler, K. (1995). *Einführung in die Phonetik des Deutschen.* Erich-Schmidt Verlag: Berlin.

Ladd, R.D. (1984). Declination: A review and some hypothesis. *Phonology Yearbook, 1,* 53–74.

Ladd, R.D. (1993). On the theoretical status of "the baseline" in modelling intonation. *Language and Speech,* 36(4), 435–451.

Ladd, R.D. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Lecuit, V., & Demolin, D. (1998). The relationship between intensity and subglottal pressure with controlled pitch. *Proceedings of the 5th International Conference on Spoken Language Processing, Sydney:* 3083–3086.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.). *Language Sound Structure* (pp. 157–233). Cambridge, MA: MIT Press.

Lieberman, P. (1967). *Intonation, perception and language.* Cambridge, MA: MIT Press.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985). Measures of the sentence intonation in read and spontaneous speech in American English. *Journal of the Acoustical Society of America, 77,* 649–657.

Löfqvist, A., Carlborg, B., & Kitzing, P. (1982). Initial validation of an indirect measure of subglottal pressure during vowels. *Journal of the Acoustical Society of America, 72(2),* 633–635.

Maeda, S. (1976). *A characterization of American English intonation.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Mücke, D., Grice, M. & Cho, T. (2014). More than a magic moment – Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics, 44,* 1–7.

Horne, M., Frid, J. & Roll, M. (2006). Timing restrictions on prosodic phrasing. In *Nordic Prosody IX,* 117–126.

Nakajima, S., & Allen, J.F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica, 50,* 197–210.

Ohala, J. J. (1978). The production of tone. In V. A. Fromkin (ed.) *Tone: A Linguistic Survey* (pp. 5–39). New York: Academic Press.

Ohala, J.J., Dunn, A. & Sprouse, R. (2004). Prosody and phonology. In *Proceedings of Speech Prosody,* Nara, Japan.

Oppenrieder, W. (1988). Intonatorische Kennzeichnung von Satzmodi. In H. Altmann (ed.), *Intonationsforschungen* (pp. 169–205). Tübingen: Niemeyer.

Parameswaran, K., Todd, D.C., & Soth, M. (2006). Altered respiratory physiology in obesity. *Canadian Respiratory Journal,* 13(4), 203–210.

Petrone, C. and Truckenbrodt, H. (in press). On falling plateaux before a prosodic boundary in German. *Journal of the International Phonetic Association*.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America, 66(2),* 363–369.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge: MIT Press.

Pike, K.L. (1945). *The intonation of American English*. University of Michigan Publications in Linguistics 1. Ann Arbor, MI: University of Michigan Press.

Pitrelli, J.F., Beckman, M.E., Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of ICSLP*, 123–126.

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics, 24,* 445–473.

Quinlan, J.R. (1993). *Programs for machine learning*. Morgan Kaufmann: San Mateo.

R Development Core Team (2012). R: A language and environment for statistical computing [version 2.15]. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Reichel, U.D. (2012). PermA and Balloon: Tools for string alignment and text processing. *Proceedings of Interspeech*, Portland, paper 346.

Rochet-Capellan, A. & Fuchs, S. (2013). Changes in breathing while listening to read speech: the effect of reader and speech mode. *Frontiers in Psychology* 4. 960.

Rochet-Capellan, A. & Fuchs, S. (2013b). The interplay of linguistic structure and breathing in German spontaneous speech. In *Proceedings of Interspeech*, paper 1228.

Schmid, C., Gendrot, C., & Adda-Decker, M. (2012). Une comparaison de la déclinaison de F0 entre le français et l'allemand journalistiques. *Proceedings of the Joint Conference JEP-TALN-RECITAL, 1,* 329–336.

Slifka, J. (2003). Respiratory constraints on speech production: Starting an utterance. *Journal of the Acoustical Society of America, 114(6)*, 3343–3353.

Stathopoulos, E.T., & Weismer, G. (1985). Oral airflow and air pressure during speech production: A comparative study of children, youths and adults. *Folia phoniatrica, 37*, 152–159.

Strik, H. (1994). *Physiological control and behaviour of the voice source in the production of prosody*. Doctoral dissertation, Catholic University Nijmegen.

Strik, H., & Boves, L. (1995). Downtrend in f0 and Psub. *Journal of Phonetics, 23,* 203–220.

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of ICSLP*, Volume 3 (pp. 1501–1504). Philadelphia, PA: ICSLP.

Swets, B., Jacovina, M.E. & Gerrig, R.J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, *50*(1), 23-51.

't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody.* Cambridge: Cambridge University Press.

Thorsen, N.G. (1980). A study of the perception of sentence intonation: Evidence from Danish. *Journal of the Acoustical Society of America, 67(3),* 1014–1030.

Thorsen, N.G. (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America*, 77, 1205–1216.

Titze, I.R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America, 85(2),* 901–906.

Titze, I.R., & Durham, P.L. (1987). Passive mechanisms influencing fundamental frequency control. In T. Baer, C. Sasaki, and K. S. Harris (eds.) *Laryngeal function in phonation and respiration* (pp. 304–319). Boston, MA: Little Brown.

van den Berg, J.W. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research, 1,* 227–244.

Van Heuven, Vincent J. & Haan, J. (2002). Temporal distribution of interrogativity markers in Dutch: a perceptual study. In C. Gussenhoven, N. Warner (eds.) *Papers in Laboratory Phonology 7* (pp. 61-86). Mouton de Gruyter, Berlin/New York.

Venables, W. N. & Ripley, B.D. (1999). *Modern Applied Statistics with S-PLUS*. Statistics and Computing Series. New York: Springer-Verlag.

Wagner, P., Tamburini, F., & Windmann, A. (2012). Objective, subjective and linguistic roads to perceptual prominence. How are they compared and why? *Proceedings of Interspeech.* Available at http://pub.uni-bielefeld.de/luur/download?func=download-File&recordOId=2529755&fileOId=2562176

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*, 423-440.

Watson, P.J., Ciccia, A.H., & Weismer, G. (2003). The relation of lung volume initiation to selected properties of speech. *Journal of the Acoustical Society of America, 113(5),* 2812–2819.

Whalen, D.H. & Kinsella-Shaw, J.M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica, 54,* 138-152.

Wiese, R. (2000). *The Phonology of German*. Oxford University Press.

Yoshida, K. (2011). Contrast maximization in f0 declination: Japanese shiki-accent dialects. *Proceedings of ICPhS XVII Hong Kong*, pp. 2228–2231.

Yuan, J., & Liberman, M. (2010). F0 declination in English and Mandarin broadcast news speech. *Proceedings of Interspeech*, 134–137.

Figure captions

Figure 1 A. Experimental set-up B. Example of signals (acoustics - top row, rib cage displacement - middle row and f0 - bottom row). (B) Examples for the reading the texts: "Diese Güte wurde später von der Maus belohnt" (Translation: *This kindness was later on rewarded by the mouse)* [13 syllables, 7 voiceless obstruents] "so unwahrscheinlich es zunächst klingt" (*so unlikely as it may sound*) [9 syllables, 9 obstruents], "Eines Tages fing sich der Löwe in einem Netz (*One day the lion was caught up in a net*) [13 syllables, 6 obstruents], "das als Falle aufgestellt war" (*which was positioned as a trap)* [8 syllables, 7 obstruents]. Superimposed lines on the RC displacement and f0 represent the linear fits used to evaluate rib cage exhalation slope and f0 declination respectively.

Figure 2. Correlation between number of voiceless obstruents and number of syllables within IPs for the read speech (left) and spontaneous speech data (right). The straight lines indicate the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.

Figure 3. Relation between f0 midline slopes (y-axis) and rib cage slopes (x-axis) for the read (left) and spontaneous (right) speech data. In each panel, data are split by the direction of the f0 midline slope.

Figures 4. Relation between f0 midline slope and number of syllables within the IPs for the read (left) and spontaneous (right) datasets. The straight line indicates the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.

Figures 5. Relation between rib cage slope and number of obstruents within the IPs for the read (left) and spontaneous (right) datasets. The straight line indicates the best fit obtained through linear regression. The grey area around the trendline marks the 99% confidence interval.