HAL
open science

The interplay of linguistic structure and breathing in German spontaneous speech

Amélie Rochet-Capellan, Susanne Fuchs

▶ **To cite this version:**

The interplay of linguistic structure and breathing in German spontaneous speech

Amélie Rochet-Capellan[1] & Susanne Fuchs[2]

[1]GIPSA-lab, UMR 5216 CNRS/Université de Grenoble –France

[2]Centre for General Linguistics, Berlin – Germany

```
Amelie.rochet-capellan@gipsa-lab.grenoble-inp.fr
fuchs@zas.gwz-berlin.de
```

Abstract

This paper investigates the relation between the linguistic structure of the breath group and breathing kinematics in spontaneous speech. 26 female speakers of German were recorded by means of an Inductance Plethysmograph. The breath group was defined as the interval of speech produced on a single exhalation. For each group several linguistic parameters (number and type of clauses, number of syllables, hesitations) were measured and the associated inhalation was characterized. The average duration of the breath group was ~3.5 s. Most of the breath groups consisted of 1-3 clauses; ~53% started with a matrix clause; ~24% with an embedded clause and ~23% with an incomplete clause (continuation, repetition, hesitation). The inhalation depth and duration varied as a function of the first clause type and with respect to the breath group length, showing some interplay between speech-planning and breathing control. Vocalized hesitations were speaker-specific and came with deeper inhalation. These results are informative for a better understanding of the interplay of speech-planning and breathing control in spontaneous speech. The findings are also relevant for applications in speech therapies and technologies.

Index Terms: spontaneous speech, breathing kinematics, breath group, inhalation pauses, syntactic clause, hesitation

1. Introduction

On a time-scale of several seconds, speech production is a sequence of short inhalations pauses followed by long exhalations with phonation. The interval of speech produced on a single exhalation is commonly defined as the breath group. It relies on linguistic, communicative and physiological constraints. The breath group is also an important unit for prosody and speech perception [1]. The present paper analyses the breath group in German spontaneous speech with respect to two main questions: (1) What is the linguistic structure of the breath group? (2) Is this structure anticipated during inhalation?

The relation of the inhalation depth and duration to the linguistic structure of the upcoming breath group reflects the interplay of speech-planning with ventilation [2-8]. These relations have been investigated in both read and spontaneous speech. These studies involved different speech tasks (e.g. sentences and texts reading, spontaneous speech with different cognitive load) and estimated breathing parameters with different methods (detection of breath noises, e.g. [2,9]; measurement of the air flow from the mouth and nose, e.g. [10]; monitoring of the kinematics of the chest wall, e.g. [3-8, 11-16], see also [17] for a comparison between acoustic and kinematic methods).

Several studies show an anticipation of the breath group length during the preceding inhalation for sentence and text reading. The inhalation depth and duration increase with the sentence length [6, 11-14, 16]. Furthermore, inhalations in sentence reading are not clearly related to the syntactic complexity (number of clauses) of the upcoming breath group [12-13]. In text reading, almost 100 % of the inhalation pauses occurs at syntactic boundaries, indicated by punctuation marks or conjunctions (e.g. *and*). These results show that the breath groups are syntactically structured [2-6, 8-10, 15]. In text reading, the inhalation depth and duration also differed with respect to syntactic marks (e.g. paragraph > period > comma) [5-6].

In spontaneous speech, the breathing pauses are not only governed by syntax but also by the cognitive processing required to generate the linguistics content [2,4,7-8,15]. This process introduces disfluencies in the speech flow. In spontaneous speech about 80% of the breathing pauses occur at syntactic constituents; the average amplitude and duration of inhalation are similar to text reading and are reflecting the length of the upcoming breath group. The average duration of breath groups is also longer than in text reading [see: 4, 7, 15, 18-19]. The ranges of variability of these parameters are larger in spontaneous speech as compared to text reading. Spontaneous speech is also characterized by the production of vocalized hesitations (*uh, um*) that have been assumed to have different functions and have been related to breathing [20, 21].

This paper evaluates the relationship between the kinematics of breathing and the linguistic structure of the breath group in German spontaneous speech. As in previous studies we consider the syntactic structure (number of clauses) and the number of syllables in the breath group. We also analyzed the type of clauses (matrix, embedded clause) and disfluencies (hesitations – *uh, um*, repetition, repairs...) in the breath group. The type of the first clause (matrix clause or embedded clause) in the breath group is an indicator of the location of inhalation relative to the linguistic structure. The association of breathing to disfluencies, and especially vocalized hesitations, is informative about the cognitive process involved in speech planning.

2. Experiment

2.1. Subjects

The participants were 26 female, native speakers of German (age: 25 years (mean) ±3.1 (standard deviation), body mass

index 21.5 ±2.1). All participants had no known history of speech, language or hearing disorders.

2.2. Experimental settings and procedure

Participants were standing up in front of a directional microphone and two loudspeakers (Figure 1.A). The spontaneous speech task was part of larger experimental protocol. After a short recording of breathing at rest and short reading, participants were instructed to listen attentively to the audio recordings of ten brief texts (151±22.1 syllables), read by a male or a female native speaker of German. The tracks were played back through the loudspeakers. After listening to each text, participants briefly summarized the story. In order to limit the movements that could interfere with the monitoring of breathing kinematics, participants were instructed to keep their hands along their trunk. Vital capacity (VC) maneuvers were run at the end of the procedure to estimate the displacement of the rib cage and the abdomen induced by VC. To do so, subjects exhaled as much air as they could and then inhaled as much air as they could.
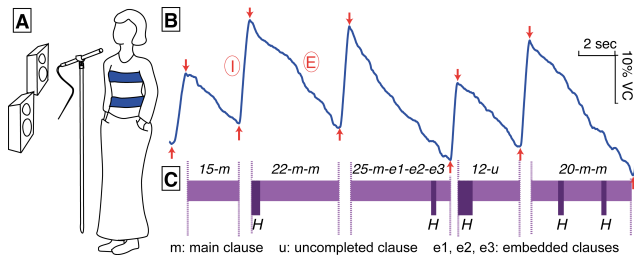

Figure 1: (A) Experimental set-up; (B) Sample breathing kinematics with inhalation (I) and exhalation phase (E). (C) Labeling of the breath groups, with number of syllables and clauses. *H* indicates the vocalized hesitations parts, see text for details.

2.3. Data acquisition, processing and labeling

The rib cage and the abdominal kinematics were recorded by means of an Inductance Plethysmograph (Respitrace™). One band was positioned at the level of the axilla (rib cage) and the other band at the level of the umbilicus (abdomen, see Figure 1.A). The acoustic and the breathing signals were recorded synchronously by means of a six channels voltage data acquisition system. The gains were the same for the thorax and the abdomen and for all the participants. All signals were sampled at 11030 Hz,

After the recording, the breathing data were sub-sampled at 200 Hz and pass-band filtered [1-40Hz]. The contribution of the rib cage and the abdomen to speech breathing varied according to the speaker. For some speakers, breathing cycles were not clear for the abdomen. For these reasons, we analyzed the sum of the rib cage and the abdomen displacements. As Respitrace™ was not calibrated, our measures could over- or sub-estimate the contribution of the thorax relative to the contribution of the abdomen to lung volume and should not be considered as a direct estimation of lung volume [22-23]. To allow comparison between speakers and conditions, displacements were expressed for each subject in %MD (Maximal Displacement). MD was the displacement corresponding to the excursion of the rib cage and the abdomen during the VC maneuver. The onset and offset of inhalations were automatically detected on the breathing signal using the velocity profiles and zero crossing. The detection was then visualized and corrected when required. The breathing cycle

was divided into an inhalation and an exhalation phase (Figure 1.B).

Speech productions were labeled in Praat [24] by detecting the onset and offset of vocalizations and by transcribing the spoken text for each breath group. The vocalized hesitations (e.g. *uh*, *um*) and the non-breathing pauses were distinguished (see Figure 1.C). On the basis of this transcription, the number of syllables was derived automatically from the output of the BALLOON toolkit [25]. The syntactic labeling of the breath groups was done by a trained phonetician. The clauses were marked by distinguishing between matrix and embedded clauses. German is a language where the position of the auxiliary verb (verb second or verb final) defines the type of clause. Mainly, the clauses with a verb in a second position were considered as matrix (also called main) clauses and those with a verb final position were considered as embedded clauses. For instance, *m-e1-e2* characterized a breath group that included one matrix clause followed by two embedded clauses, with the first one (*e1*) referring to the matrix clause (*m*), and the second one (*e2*) referring to the first embedded clause (*e1*), see Figure 1.C. The third category, uncompleted clauses (*u*), included words or groups of words corresponding to hesitations, repetitions or repairs.

2.4. Data selection

Our data set included 1467 breath groups. We discarded 45 groups that were perturbed by laugh, cough or body movements. The number of clauses ranged from 1 to 7 (2.11 (mean) ±1.13 (standard error)). The dataset was restricted to groups with 1-3 clauses. They represented 88% of the observations and were produced by all subjects. Only groups starting with *m*, *e1*, *e2* or *u* were considered in this study (99% of the groups with 1-3 clauses).

2.5. Measures and analyses

We estimated: (1) the duration of the breath group (*dur_g*), as the time interval from speech onset to speech offset; (2) the amplitude (*amp_I*) and duration (*dur_I*) of inhalation; (3) the relationship between *amp_I* and the amplitude of exhalation (*amp_IE*, *amp_I* divided by the amplitude of exhalation). This last measure evaluates if speakers exhale more air (*amp_IE* < 1), less air (*amp_IE* > 1) or the same amount of air (*amp_IE* = 1) than they have just inhaled to produce the breath group. This measure could not be taken as an indicator of the reserve volume consumption, as displacements values were not expressed relative to a zero volume.

We considered four main factors: (1) the number of clauses in the breath group (*n_clauses, 1*, *2*, *3*); (2) the number of syllables *n_syll* (continuous factor); (3) the type of first clause *f_clause* (*m, e1, e2, u*); (4) the type of hesitation: *t_hesi* (levels: *none*, at least one at onset: *onset*, at least one not at onset: *elsewhere*).

The effects of *n_syll*, *n_clauses* and *f_clause* on the different parameters were tested as fixed factors effects using Linear Mixed Models (LMM), with *subject* as a random factor. The interactions between factors were not significant and therefore, additive models were calculated. For *dur_I* and *amp_IE* the log values were used to satisfy normality. An analysis of hesitation was introduced in a second step with *subject* as random factor and *n_syll* and *t_hesi* as fixed factors. All the effects reported significant were satisfying the criteria pMCMC <.01.

3. Results

Table I. Description of the breath groups according to the number of clauses and to the type of the first clause. *NB:* Number of breath groups; *n_syll*: Average number of syllables; *dur_g*: average duration (± one standard error).

		F_clause				
		m	*e1*	*e2*	*u*	*All*
1	*NB.*	218	84	40	160	502
	n_syll	11.7 (±.35)	11.3 (±.59)	11.2 (±.66)	6.2 (±.32)	9.9 (±.24)
	dur_g	2.61 (±.08)	2.57 (±.16)	2.57 (±.17)	1.53 (±.07)	2.26 (±.09)
2	*NB.*	272	77	29	82	460
	n_syll	17.7 (±.36)	18.2 (±.56)	18.3 (±.91)	13.8 (±.54)	17.1 (±.27)
	dur_g	3.95 (±.08)	3.90 (±.15)	3.88 (±.35)	3.27 (±.15)	3.82 (±.07)
3	*NB.*	171	40	14	46	271
	n_syll	26.4 (±.47)	26.3 (±1.12)	22.1(±1.74)	21.0 (±.84)	25.3 (±.40)
	dur_g	5.61 (±.12)	5.30 (±.18)	4.50 (±.45)	4.68 (±.22)	5.35 (±.09)
All	*NB.*	661	201	83	288	1233
	n_syll	17.9 (±.31)	17.0 (±.56)	15.5 (±.77)	10.8 (±.42)	15.9 (±.24)
	dur_g	3.94 (±.07)	3.62 (±.12)	3.35 (±.19)	2.53 (±.10)	3.52 (±.05)

(row label: **N_clauses**)

3.1. Linguistic structure of the breath group

The average characteristics of the breath groups and their repartition according to the first clause and to the number of clauses are displayed in Table 1. Speakers produced from 13 to 99 breath groups (47.4 (mean) ±4.5 (sterr), Figure 2). Half of the breath groups (53%) started with a matrix clause *(m)*, a quarter (24%) with and embedded clause and the last quarter (23%) with an uncompleted clause (*u*). On average, the breath group included ~15.9 syllables (range: 1 to 50), and lasted ~3.5 s (range: .17 to 12.1). The number of syllables and the duration of the groups significantly increased with the number of clauses (~+7.5 syllables and +1.5 s per supplementary clause), but were similar for groups starting with a matrix as compared to an embedded clause. Groups starting with an uncompleted clause were ~6 syllables and 1.1 s shorter than the other groups.
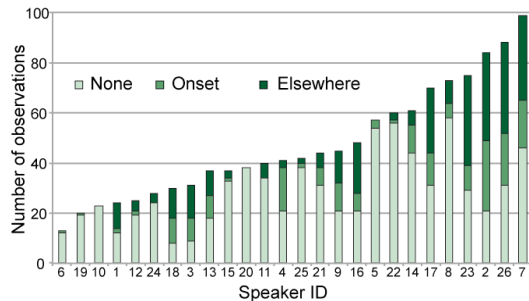

Figure 2. Number of breath groups for each speaker with repartition of groups in: no hesitation (none), at least one hesitation at onset or elsewhere

The percentage of breath groups with vocalized hesitations ranged form 0 to more than 50% according to the subject (average 40%, see Figure 2). Among the breath groups with at least one hesitation (n=482), 40% started with a hesitation. Note that the groups with at least one hesitation not at the onset of the group were longer than the groups starting with a hesitation (~+3syllables and ~+749 ms) and than the groups without hesitation (~+3syllables and ~+1246 ms). The effect of hesitation type (*t_hesi*) on the number of syllables and the duration of the group were significant but didn't interact with the effect of the first clause.
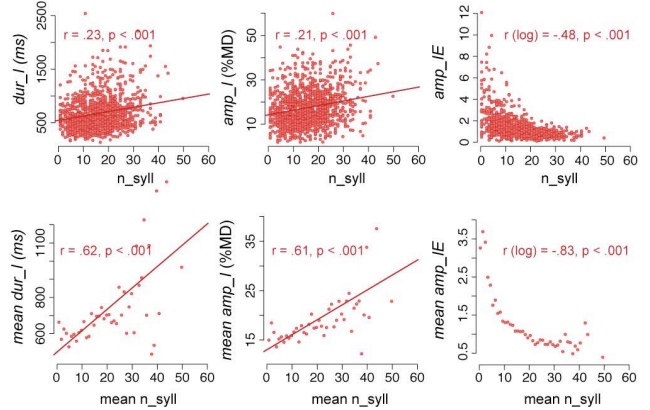

Figure 3. Correlations between *n_syll* in the breath group with: *dur_I* and *amp_I* and *amp_IE,* all values (top), average (bottom). Correlations values for *amp_IE* are indicated for *log(amp_IE), see text for details*.

3.2. Breathing kinematics

On average, the duration of inhalation was 676 ms (±8.5) and the amplitude was 17.6 %MD (± 0.2). The amplitude and the duration of inhalation depended both on the length of the breath group and on the type of the first clause. These values were also positively correlated with the number of syllables (r = ~.20 for all values and r = ~.60 for average correlations, see Figure 3, first two columns). LMM showed a significant effect of *n_syll* on both *amp_I* and *dur_I*.
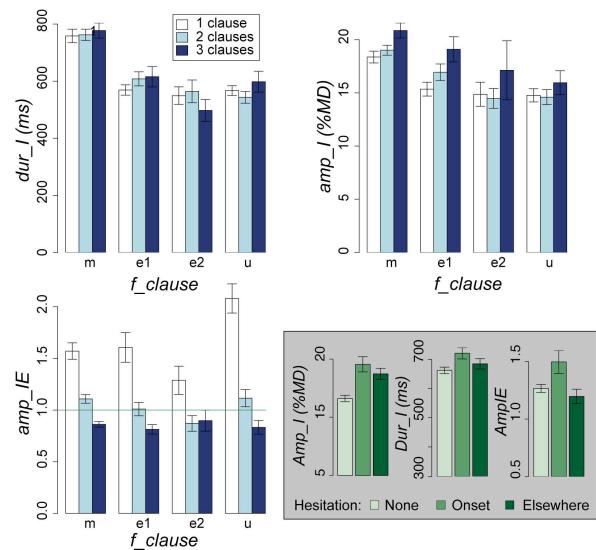

Figure 4. Average and standard errors of *dur_I*, *amp_I* and *amp_IE* according to *n_clauses* and *f_clause* (white panels) and to the type of hesitation in the breath group (gray panel)

The duration of inhalation (Figure 4.A) significantly increased from 1 to 2 (+26 ms) and 2 to 3 (+36 ms) clauses. *Dur_I* was also longer for groups starting with a matrix clause as compared to other types of clauses (+197 ms). Inhalation (Figure 4.B) was significantly deeper when the first clause of the upcoming group was a matrix clause (+3.5 %MD) than any other clauses. Yet, *amp_I* did not significantly depend on the number of clauses.

The analysis of the inhalation displacement relative to the exhalation displacement (*amp_IE*, Figure 3 and 4) shows: (1) that *amp_IE* was close to 1 for groups with 2 clauses and groups with 15-18 syllables; (2) a significant linear correlation between the logarithm of *amp_IE* with the number of syllables (-.48 for all values, -.83 for average, significant effect of *n_syll*); (3) an effect of the number of clauses (*1 > 2 > 3*); (4) no significant effect of the type of the first clause. Hence, on average, the inhalation displacement was similar to the exhalation displacement for groups with 2 clauses or 15-18 syllables, larger for shorter groups and smaller for longer groups.

Inhalations were deeper (+2.54 %MD) and longer (+41 ms) for the breath groups with at least one hesitation as compare with no hesitation (Figure 4). The effect of *t_hesi* on *amp_IE* was not significant when the number of syllables was taken into account.

4. Discussion

The present study investigated the linguistic structure of the breath group in German spontaneous speech and evaluated if this structure is reflected in breathing kinematics. The important findings are:

(1) Inhalations occur at syntactic boundaries (before a matrix or an embedded clause) or before a disfluency (uncompleted clause, repetition, hesitation, repair);

(2) Inhalation depth and duration reflect: (2.1) the length of the breath group (number of syllables); (2.2) the type of the first clause, with deeper and longer inhalation for groups starting with a matrix clause as compared to the other groups; (2.3) vocalized hesitations, with deeper and longer inhalations for groups that include at least one vocalized hesitation as compared to none;

(3) Syntactic complexity (number of clauses) is reflected only in the duration but not in the amplitude of inhalation;

(4) On average the amplitude of exhalation is similar to the amplitude of inhalation for groups with 2 clauses or 15-18 syllables.

The observation that most of the inhalation pauses respect the syntactic organization of speech is consistent with previous work on English spontaneous speech [7,15]. The average duration (3.5 s), the number of syllables in the breath group (16 syllables) and the duration of inhalation (~.7 s) are also similar to values reported in the literature on English language ([7,8,15]).

As described in the introduction, previous studies found deeper and longer inhalations for longer utterances. Our dataset also show these relations. However, we also found that inhalations were deeper and longer for the breath groups starting with a matrix clause and for the groups including hesitations as compared to the other groups. To our knowledge, the relationship between the type of the first clause and hesitation to inhalation parameters have not been investigated so far for spontaneous speech. This relation is important with respect to the understanding of speech planning. It suggests that speaker inhale more air: (1) when they are starting a matrix clause that may come with other related clauses; (2) when they produce hesitations and do not know exactly what they are going to say. In this case, they can use vocalized hesitations as fillers during the exhalation phase, which could help to preserve ventilation and speech at the same time [21]. The fact that the breath groups with a hesitation at the onset were shorter than groups with a

later hesitation shows that when hesitation came at the onset of the group, speaker probably inhaled again soon after it.

We also found that groups with an average number of syllables (15-18) show similar exhalation and inhalation amplitudes. These breath groups correspond to 2 clauses and could be a "favored" association between linguistic structure and breathing. This hypothesis should be tested by considering inter-speaker variability and speaker-specific lung volume capacities.

The speech task used in the present study required speakers to summarize the story they have just heard. This task is cognitively demanding and could have influenced the production of hesitations and the breathing profiles. This is in line with inter-speakers variability we found with respect to the number of breath groups and hesitations produced in the current task. To our knowledge only [8] have investigated the possible effect of cognitive load on breathing kinematics during spontaneous speech. We think it is important to distinguish between speaker-specific behaviors according to the task (e.g. variation in disfluency, hesitations).

5. Limits and perspectives

This study is a first analysis of a larger corpus of breathing kinematics in German spontaneous speech that now includes more than 50 speakers. Our global aim is to understand the interplay of speech planning and breathing in unconstrained speech. From the current study some first issues appear: (1) it is difficult to distinguish between the effect of the number of syllables and the effect of the number of clauses. Note that the quartile of the average number of syllables (10-15-21) were close to the average number of syllables in 1, 2, and 3 clauses, respectively (10-17-25 syllables); (2) Uncompleted clauses should be analyzed in more detail by splitting between hesitations, repairs and repetitions, that could have specific effect on breathing; (3) the amplitude of inhalation anticipates the upcoming breath group, but may also rely on what happened before [9]. This may be especially true for groups starting with an embedded clause. The next step is also to characterize the breath group in spontaneous speech not only as an individual unit but as a temporal sequence that depends on the preceding and following speech.

Speaker-specific behavior and context effects should also be considered. Previous studies on read and spontaneous speech, found that the properties of the breath group and their relations to inhalation parameters are speaker-specific [10,13], varied with age [11], cognitive load [8], speech rate [3] and loudness [16,19]. A large variability has also been observed for a same subject across repetitions and according to her emotional state [6-7, 10]. The sensitivity of speakers' breathing regarding these multiple influences is important to understand the interplay between linguistics and respiration and may provide a fundamental tool for pathological diagnostics and speech therapy. Furthermore, implementing breathing in speech synthesis may improve the naturalness of speech synthesizers.

6. Acknowledgements

7. References

[1] Lieberman, P., Intonation, Perception and Language. (1967), Cambridge MA: MIT Press.

[2] Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). Temporal patterns of cognitive activity and breath control in speech. Lang Speech, 8, 236--242.

[3] Grosjean, F. & Collins, M. (1979). Breathing, pausing and reading. Phonetica, 36(2), 98--114.

[4] Conrad, B. & Schonle, P. (1979). Speech and respiration. Arch Psychiatr Nervenkr, 226, 251--268.

[5] Conrad, B., Thalacker, S., & Schonle, P. (1983). Speech respiration as an indicator of integrative contextual processing. Folia Phoniatr (Basel), 35, 220--225.

[6] Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: lung volumes, speech intensity, and linguistic factors. J Speech Hear Res, 37, 535--556.

[7] Winkworth, A. L., Davis, P. J., Adams, R. D., & Ellis, E. (1995). Breathing patterns during spontaneous speech. J Speech Hear Res, 38, 124--144.

[8] Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. J Speech Hear Res, 39, 93--104.

[9] Bailly, G. and Gouvernayre, C. (2001). Pauses and respiratory markers of the structure of book reading. in Interspeech. 2012. Portland, OR.

[10] Teston, B. and Autesserre, D. (1987). L' aérodynamique du souffle phonatoire utilisé dans la lecture d'un texte en français. in International Congress of Phonetic Sciences (ICPhS). Estonia, University of Tallin. p. 33-36.

[11] Sperry, E. E. & Klich, R. J. (1992). Speech breathing in senescent and younger women during oral reading. J Speech Hear Res, 35, 1246--1255.

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

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

[14] McFarland, D. H. & Smith, A. (1992). Effects of vocal task and respiratory phase on prephonatory chest wall movements. J Speech Hear Res, 35, 971--982.

[15] Wang, Y. T., Green, J. R., Nip, I. S., Kent, R. D., & Kent, J. F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. Folia Phoniatr Logop, 62, 297--302.

[16] Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults}. Respir Physiol Neurobiol, 164, 323--330.

[17] Wang, Y.T., Nip, I.S.B., Green, J.R., Kent, R.D., Kent, J.F., Ullman, C. Accuracy of perceptual and acoustic methods for the detection of inspiratory loci in spontaneous speech. Behavior Research Methods, 2012. 44(4): p. 1121-1128

[18] McFarland, D. H. (2001). Respiratory markers of conversational interaction}. J. Speech Lang. Hear. Res., 44, 128

[19] Huber, J. E. (2007). Effect of cues to increase sound pressure level on respiratory kinematic patterns during connected speech}. J. Speech Lang. Hear. Res., 50, 621--634.

[20] Ferreira, F. and Bailey K. G.D. (2004). Disfluencies and human language comprehension. Trends in Cognitive Sciences, 8(5), 231–237.

[21] Schonle, P. W. & Conrad, B. (1985). Hesitation vowels: a motor speech respiration hypothesis. Neurosci. Lett., 55, 293--296.

[22] 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.

[23] Banzett, R. B., Mahan, S. T., Garner, D. M., Brughera, A. & Loring, S. H. (1995). A simple and reliable method to calibrate respiratory magnetometers and Respitrace. Journal of Applied Physiology, 79(6), 2169-2176.

[24] Boersma, P. and D. Weenink, Praat, a System for doing Phonetics by Computer, version 3.4, in Institute of Phonetic Sciences of the University of Amsterdam, Report 132. 182 pages. 1996.

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