INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Temporal coordination of articulatory and respiratory events prior to speech initiation

Oksana Rasskazova[1 2]*, Christine Mooshammer*[1]*, Susanne Fuchs*[2]

[1]Humboldt-Universität zu Berlin, Germany
[2]Leibniz-Centre General Linguistics, Germany

rasskazova@leibniz-zas.de, mooshamc@hu-berlin.de, fuchs@leibniz-zas.de

Abstract

The investigation of the speech planning processes, in particular the timing between acoustic and articulatory onset, has recently received a lot of attention. Respiration has not been considered in this process so far, although it is involved and may be well coordinated with the oral articulators prior and at the onset of the utterance. In light of these considerations, we investigated the temporal coordination between acoustic, respiratory and articulatory events prior to utterance onset. For this purpose 12 native speakers of German have been recorded with Electromagnetic Articulography and Inductance Plethysmography reading sentences that were controlled for length and stress of the first word. The initial segment of the utterance was either /t/ or /n/. The results for six speakers so far indicate that early speech preparation consists of mouth opening during the inhalation phase. The onset of expiration seems to be tightly coupled with the acoustic and the articulatory onset, particularly with the constriction interval of the tongue tip gesture in the first segment. Manner of articulation of the initial segment seems to affect the temporal fine-tuning of preparatory events.

Index Terms: oral-respiratory coordination, articulatory speech planning, pauses, Electromagnetic Articulography, respiration

1. Introduction

Pauses are acoustically silent intervals during which speakers plan subsequent utterances on different levels [1]. Planning on the phonetic level involves articulatory movements towards the first segments. Respiration, in particular inhalation depth, might be involved as well, but is related to a longer time span, the length of the upcoming utterance, e.g. [2], [3], [4], [5]. How both are coordinated with another, is to a large extent unclear, even if previous work mentioned a tight link between the acoustic onset of speech and exhalation onset [6], [7]. So far, most attention has focused on what is happening in the vocal tract during silent pauses. No matter what researchers found in detail, the articulators are far from immobile during the silent phase, they move, even if this motion is not audible and may not always be target specific.

Previous studies mentioned speaker-specific differences in articulatory behavior during silent pauses [8]. Various types of tongue postures have been described and termed speech-ready posture or tongue rest position e.g. [9], [10], [2], [11], [12]. The lips also vary in their position. During pauses they might be open, half open, closed or might be in a transition between those states. That means, various vocal tract activities occur during speech preparation, which seem to be temporally organized with each other. The following questions arise:

1.) What may influence articulatory behavior before starting an utterance? Is the articulatory motion of oral articulators a by-product of inhalation and the opening of the vocal tract passage or is it a preparatory motion for an upcoming utterance?
2.) How are these different articulatory motions coordinated?
3.) What is the stability in the coordination of these preparatory events?
4.) To what extent are these articulatory motions affected by the initial segment?

Theoretically, our investigation is closely linked to Articulatory Phonology [13], and specifically Task dynamics [14]. Task dynamics defines functional units that aim towards specific goals in order to produce speech. These goals are reached by the coordinated actions of different articulators. However, the temporal organization of vocal tract activities prior the speech initiation have not been investigated in detail, even if the initial position from which articulators start to move, is crucial for modelling and is associated with a "neutral attractor" [15].

The results of previous studies on temporal aspects of speech planning show that the movements of the articulators usually start before the acoustic onset. Depending on the manner of articulation of the initial segment, this delay is roughly 120 -180 ms [16], [17], [12]. The implications for such preparatory motions are manifold. They have been discussed with respect to methodological issues in reaction time experiments, e.g. in [18],[16],[19], [20], using EEG and fMRI [21]. In such experiments the onset of speech is often considered as a crucial reference point defined on the basis of the acoustic signal ("voicekey") if though the articulatory motions start before it is detectable in the acoustic signal.

Another factor that has been largely ignored is that inhalation occurs before utterance onset and is involved in the speech planning process. The respiratory activities may play a crucial role in the temporal organization of the motor plan. Indications for this relationship are coming from studies on respiratory activities prior to the utterance begin [22], [5]. There is a correlation between the actions of the respiratory system and the length of the upcoming sentence, such as longer inhalation duration as well deeper inhalations are found prior to longer upcoming sentences [5]. The peak of the subglottal pressure also correlates with segmental and prosodic properties when starting an utterance [22].

To our knowledge, respiration has not yet been considered as a functional unit similar to other articulators in Task dynamics, although breathing activities cannot be separated from articulation (with a few exceptions where sounds are non-pulmonic). Furthermore, speech production takes place on the expiratory air stream. Respiration is, however, slower than oral articulation, but both should be well timed with each other, particularly at the onset of speech.

The aim of the present study is to investigate the coordination of respiratory, acoustic and articulation events prior to the utterance begin. Thus, inhalation onset might be tightly coupled

http://dx.doi.org/10.21437/Interspeech.2019-2876

with lip opening. Since speech production takes place on the expiratory air-stream, our working hypothesis is that exhalation onset will be tightly coupled with articulatory movements and acoustic onset of the first segment of an upcoming utterance. Furthermore, we explore if there is a coordination between exhalation onset and a specific phase of the articulatory gesture of an upcoming segment.

2. Methods and Results

2.1. Methods

2.1.1. Participants and Material

Twelve native German speakers, aged between 22 and 38 years old, without known history of respiratory or articulatory disorders and hearing impairment participated in the study. The participants performed a reading task. The speech material involved eleven utterances that consisted of two sentences each. The utterances were presented in randomized order on a computer screen. They were mixed with various filler sentences, which differed in their structure and consisted of one sentence only. The target utterance was controlled for sentence length and word stress. The initial segment of the first word of the target utterances was /a/, /t/, /n/, /h/ or /ʃ/. Five repetitions of each target utterance were produced. In this paper the focus is laid on /t/ and /n/ as initial segments occurring in six utterances (30 stimuli per subject). Currently, six subjects have been analyzed. More data is currently under analysis.

2.1.2. Recording procedure

Respiration, speech kinematics and acoustics were simultaneously recorded by means of Electromagnetic Articulography (EMA (AG501)) and Inductance Plethysmography. Acoustic data were recorded at 44.1 kHz using a shotgun microphone located in front of the speakers. The EMA sensors were attached to the tongue tip (TT), tongue middle (TM) and tongue back (TB), the jaw, and the upper and lower lips (UL, LL). Four reference sensors were included to compensate for head movements. The articulatory data were recorded at a sampling rate of 1250 Hz and then downsampled to 250 Hz for calibration and post-processing in MATLAB. The data were corrected for head movement and then rotated and translated to the bite plane or to the fictional plane between the upper incisors and the nose.

To record the respiration data two elasticized bands, one around the rib-cage and another around the abdomen, were put on the participants. During the recording session participants were sitting in a straight position under the EMA, in front of the monitor. The two respiratory signals were recorded together with the audio signal on a multi-channel DAC 6B recorder. The Inductance Plethysmography system was connected to the EMA system by means of synchronization box (from Carstens Medizinelektornik). Thereby, the synchronization impulse of the EMA recordings were transferred to one channel of the multi-channel DAC recorder. The other three channels were acoustics, thoracic and abdominal volume changes. To prevent both systems from drifting apart from each other, we used the Arduino Processor (synchronisation tool implemented by Philip Hoole) that controls the starting point of each system.

In the post-processing procedure respiratory data were cut according to the synchronization impulse from the EMA system. The duration of articulatory and respiration trials was compared and corrected by calculating a cross correlation between two acoustic signals recorded by the two systems.

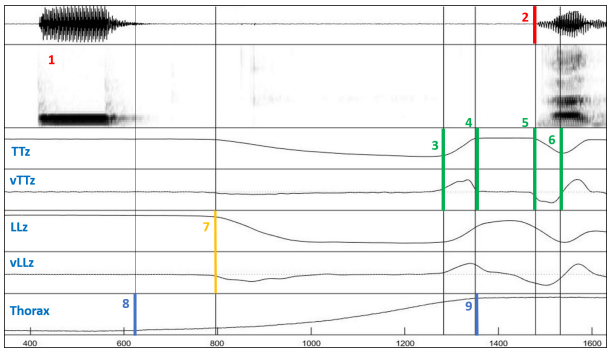

Figure 1: *Labelling procedure of measuring acoustic, articulatory and respiratory parameters during silent speech interval prior utterance begin. The acoustic signal was labelled for the onset of the beep signal (1) and the acoustic onset (2). The tongue tip (TT) signal was labelled for (3) gesture onset, (4) nucleus onset and offset (5) and gesture offset (6). The lower lip gesture (LL) was labelled for its onset (7). The respiratory signal of the thorax was labelled for the inhalation (8) and exhalation (9) onsets.*

Before each trial participants heard the beep signal. The participants received this signal as a sign that they can start reading the presented speech material. The temporal occurrence of the beep signal as well the time speakers started to read the stimuli were only recorded, but not controlled.

2.1.3. Measurements

The acoustic, articulatory and respiratory data were labelled with the visualization and labelling tool MVIEW ([23]), written in MATLAB. For our research question we analyzed the time span starting briefly before the beep signal and ending at the onset of an upcoming utterance. The following events prior to the utterance onset were analyzed: beep and acoustic onset of speech, inhalation and exhalation onset, movement onset of the lower lip as well the tongue tip gesture of the upcoming alveolar segment.

Acoustic and respiration data were analyzed using a self-written labelling procedure, which detects acoustic onsets for reaction-time data. The algorithm finds automatically the acoustic beep and the acoustic onset of the following utterance based on the RMS peak amplitudes. Temporal respiratory events were manually labelled as inhalation minima and maxima based on either thoracic or abdominal signals (cf. Figure 1). All analyzed speakers showed more pronounced thoracic than abdominal movements during reading. The movement onset of the lower lip, the gestural on- and offsets as well as the nucleus' on- and offset of the tongue tip movement were determined by automatically finding the gesture onset and offset of the target segment (/t/ and /n/) by using a 20% threshold criterion of the tangential velocity signal (cf. Figure 1).

Some data had to be partially excluded due to various reasons. Sometimes, speakers started to inhale much earlier than it can be observed in the pre-processed signals. In these cases it was not possible to determine the time point of the inhalation onset. In some cases they also did not produce the alveolar closing gesture towards the initial consonant /n/ or /t/. They maintained the tongue tip closure already at alveolar place of articulation during the whole silent interval. For these cases only the opening gesture of the tongue tip was labelled.

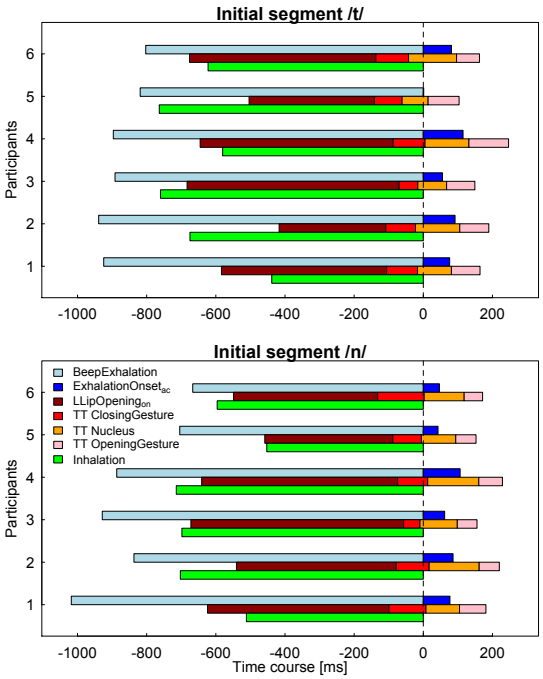

Figure 2: *Average duration delays of preparatory events for each speaker relative to the exhalation onset (dashed line) for initial /t/ (upper panel) and initial /n (lower panel)*

Statistics for the presented data were carried out in an exploratory way following work on gestural cohesion [24], [25] and by performing linear mixed effect models with the initial segment as independent variable and speaker as random factor, using R 3.5.2 [26].

2.2. Results

The lack of previous research and findings on this topic does not allow us to make clear hypotheses and test them. However, following the general notion that for longer utterances speech is produced on the egressive air-stream, our working hypothesis is that the initial segment is coordinated with the onset of exhalation in some way. Moreover, we suppose that manner of articulation of the initial segment might have an effect on the coordination between oral and respiratory events. In this section we will investigate two kinds of coordination: the lip opening motion with the onset of inhalation and second the articulatory gesture for the initial segment and the onset of exhalation.

Figure 2 shows the average duration delays for each speaker for the initial segments /t/ (left) and /n/ (right) relative to the exhalation onset, which has been set to zero. The first row of bars for each speaker shows the time between the beep and the exhalation onset (light blue) and between the exhalation onset and the acoustic onset (dark blue). The time span between the beep signal and the acoustic onset has not been controlled during experiment. Thus, the duration of silent intervals prior the utterance begin varies largely, especially among speakers. Furthermore, it can be seen that the acoustic onset happens with a short delay of approximately 69 ms after the onset of exhalation. This delay is not significantly different comparing /n/ and /t/ ($\beta = 2.1ms, t = 0.46$).

Mouth opening (dark red, in the second row), measured at the lower lip, always happened after the beep signal (light blue box). We assume that the lip opening is somehow related to inhalation through the mouth. The inhalation onset is shown in the third row as green box. As can be seen in Figure 2 there is a fair amount of speaker variability for the temporal alignment between mouth opening and inhalation onset. Speakers 2 and 9 (/t/ only) inhale before mouth opening whereas speaker 1 starts the inhalation after mouth opening. For the other speakers these two events happen at approximately the same time.

Tongue tip gesture phases and exhalation onset are shown as lighter red boxes in Figure 2. For both, /t/ and /n/, the event that is closest to exhalation onset is the onset of the constriction phase (see TT nucleus, and time mark 4 in Figure 1). For /t/ the onset of the constriction occurs on average 23 ms before the onset of exhalation, whereas for /n/ it is almost synchronous with exhalation onset (t = 5.7, p <0.001). Earlier onsets relative to the exhalation onset for /t/ than /n/ was also significant for the onset of tongue tip closure, the offset of the nucleus and the offset of opening movement (see also Figure 1, time marks 3, 5 and 6).

The latencies between respiratory and articulatory events show that the onset of exhalation is almost synchronous with the onset of the constriction phase for /n/ and /t/ (nucleus onset). In order to test whether this latency is also stable, the variability, quantified as Relative Standard Deviation (RSD), was compared with the other latencies (see Figure 3). Generally, all latencies including the exhalation onset show a relatively low variability of maximally 5 %. In contrast, the mean RSD for the latency of inhalation onset to the lower lip onset (not shown in Figure 3) is around 10 % with a maximum of 44 %. Despite the obvious inconsistent behaviour of the speakers there seems to be a tendency for larger RSDs for the tongue tip closing onset (see Figure 1, time mark 3) compared to the latency of the tongue tip nucleus on- and offset (4 and 5 in Figure 1). The acoustic onset for /n/ shows lower variability than for /t/. This might be due to the fact that for the stop the acoustic onset is the burst which is closer to the end of the tongue tip nucleus, whereas for /n/ the acoustic onset co-occurs with the tongue tip nucleus onset.

3. Discussion

In summary, we found evidence for temporal alignment between oral articulators. Movement initiation of the first segment of the utterance starts during the final phase of the inhalation. This anticipation can be interpreted as evidence for a close gestural cohesion between respiration and oral action for gestural organization.

The onset of exhalation was almost synchronous with the nucleus of the alveolar closure and showed relatively low variability. This suggests that these two events are timed with each other. In contrast, larger relative variability was found at the onset of the tongue tip closure relative to the exhalation onset.

Moreover, the timing seems to be sensitive to the identity of the initial segment. Speakers initiate the oral gestures for the nasal /n/ later (relative to the exhalation onset) than for the stop /t/. Even though /t/ and /n/ are both alveolar stops, they differ in the connection of the oral to the nasal cavity. For /t/ the velar port needs to be closed to build up pressure in the oral cavity. Velar closure may already affect respiration, because inhalation through the nose may not be possible any more, while for /n/ inhalation can be accomplished via mouth or nose. Based on our recordings, however, we cannot differentiate between mouth and nose breathing. Both affect lung volume in a similar manner. In a next step we will additionally investigate the

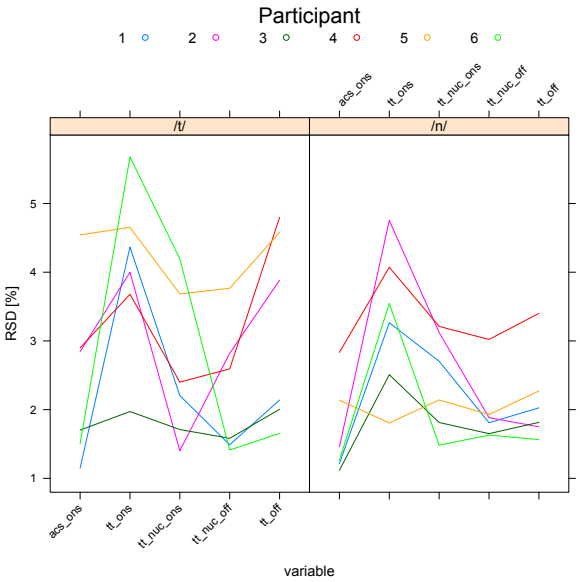

Figure 3: *Relative Standard Deviation in [%] of average duration relative to the exhalation onset (dashed line) are given. Different colours correspond to different speakers. Variability of the different delays are given on the x-axis while the amount of Relative Standard Deviation is given on the y-axis. The left graph corresponds to /t/ as an initial segment while the right graph corresponds to /n/.*

coordination between respiratory and articulatory actions in the other recorded initial segments /a, h, ʃ/. Following the presented results, we expect that the voiceless fricatives exhibit a tighter coordination between respiration and oral articulation in comparison to the low vowel.

Speaker-specific behaviour was found for inhalation onset. At this point, oral events are not yet strongly timed with respiration. Some speakers start the inhalation before they start the preparatory mouth opening, some speakers start at the same time and some start it inhalation after they open the mouth. Given these differences, we suggest that these preparatory articulatory motions are not purely a by-product of inhalation and the respective mouth opening, but an active preparation of the initial segment.

4. Conclusions

In this pilot study we have shown that the respiratory system, which works on a large time scale and has a vital function, and the oral articulatory system, which works much faster on a gestural level, seem to be interconnected and integrated in the speech planning process prior to speech initiation.

5. Acknowledgements

This research is funded by German Federal Ministry of Education and Research. We would like to thank Alina Zöllner, Megumi Terada and Jörg Dreyer for assistance with running the experiments.

6. References

[1] J. Krivokapić, "Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130397, 2014.

[2] V. Ramanarayanan, E. Bresch, D. Byrd, L. Goldstein, and S. S. Narayanan, "Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. EL160–EL165, 2009.

[3] J. M. Scobbie, S. Schaeffler, and I. Mennen, "Audible aspects of speech preparation," *Proceedings of 17th ICPhS, Hong Kong*, 2011.

[4] D. H. Whalen and J. M. Kinsella-Shaw, "Exploring the relationship of inspiration duration to utterance duration," *Phonetica*, vol. 54, no. 3-4, pp. 138–152, 1997.

[5] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, "Acoustic and respiratory evidence for utterance planning in german," *Journal of Phonetics*, vol. 41, no. 1, pp. 29–47, 2013.

[6] A. Rochet-Capellan and S. Fuchs, "Take a breath and take the turn: how breathing meets turns in spontaneous dialogue," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130399, 2014.

[7] M. Włodarczak and M. Heldner, "Respiratory constraints in verbal and non-verbal communication," *Frontiers in psychology*, vol. 8, p. 708, 2017.

[8] J. Krivokapic, W. Styler, B. Parrell, and J. Kim, "Pause postures in american english," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2584–2584, 2017.

[9] B. Gick, I. Wilson, K. Koch, and C. Cook, "Language-specific articulatory settings: Evidence from inter-utterance rest position," *Phonetica*, vol. 61, no. 4, pp. 220–233, 2004.

[10] I. Wilson and B. Gick, "Articulatory settings of french and english monolinguals and bilinguals," *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3295–3296, 2006.

[11] V. Ramanarayanan, L. Goldstein, D. Byrd, and S. S. Narayanan, "An investigation of articulatory setting using real-time magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 510–519, 2013.

[12] O. Rasskazova, C. Mooshammer, and S. Fuchs, "Articulatory settings during inter-speech pauses," *Proceedings of the Conference on Phonetics Phonology in German-speaking countries (PI&P13)*, p. 161, 2018.

[13] C. P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[14] E. L. Saltzman and K. G. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological psychology*, vol. 1, no. 4, pp. 333–382, 1989.

[15] J. Simko and F. Cummins, "Embodied task dynamics." *Psychological review*, vol. 117, no. 4, p. 1229, 2010.

[16] C. Mooshammer, L. Goldstein, H. Nam, S. McClure, E. Saltzman, and M. Tiede, "Bridging planning and execution: Temporal planning of syllables," *Journal of phonetics*, vol. 40, no. 3, pp. 374–389, 2012.

[17] P. Palo, S. Schaeffler, and J. M. Scobbie, "Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tongue ultrasound," *Proceedings of the 18th ICPhS, Glasgow*, 2015.

[18] S. Schaeffler, J. M. Scobbie, and F. Schaeffler, "Measuring reaction times: vocalisation vs. articulation," in *Proceedings of the 10th International Seminar in Speech Production (ISSP 10)*, 2014.

[19] K. Rastle, K. P. Croot, J. M. Harrington, and M. Coltheart, "Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 5, p. 1083, 2005.

[20] A. H. Kawamoto, Q. Liu, K. Mura, and A. Sanchez, "Articulatory preparation in the delayed naming task," *Journal of Memory and Language*, vol. 58, no. 2, pp. 347–365, 2008.

[21] G. Ouyang, W. Sommer, C. Zhou, S. Aristei, T. Pinkpank, and R. A. Rahman, "Articulation artifacts during overt language production in event-related brain potentials: Description and correction," *Brain topography*, vol. 29, no. 6, pp. 791–813, 2016.

[22] J. Slifka, "Respiratory constraints on speech production: Starting an utterance," *The Journal of the Acoustical Society of America*, vol. 114, no. 6, pp. 3343–3353, 2003.

[23] M. Tiede, "Mview: software for visualization and analysis of concurrently recorded movement data," *New Haven, CT: Haskins Laboratories*, 2005.

[24] C. Mooshammer, P. Hoole, and A. Geumann, "Interarticulator cohesion within coronal consonant production," *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 1028–1039, 2006. [Online]. Available: https://doi.org/10.1121/1.2208430

[25] J. Brunner, C. Geng, S. Sotiropoulou, and A. Gafos, "Timing of german onset and word boundary clusters," *Laboratory Phonology*, vol. 5, no. 4, pp. 403–454, 2014.

[26] R Core Team, "R: A language and environment for statistical computing," Wien, 2018.