Interspeech 2018
2-6 September 2018, Hyderabad

Respiratory and Respiratory Muscular Control in JL1's and JL2's Text Reading Utilizing 4-RSTs and a Soft Respiratory Mask with a Two-Way Bulb

Toshiko Isei-Jaakkola[1], *Keiko Ochi*[2], *Keikichi Hirose*[3]

[1]Chubu University, Japan
[2]Tokyo University of Technology, Japan
[3]National Institute of Informatics, Japan
tiseij@isc.chubu.ac.jp, ochikk@stf.teu.ac.jp, hirose@gavo.t.u-tokyo.ac.jp

Abstract

To investigate how respiratory muscles and respiration are controlled during L1 and L2 text readings, experiments were conducted to acquire data on (A) upper- and lower- chest and upper- and lower- abdominal movements, utilizing four respiratory strain-gauge transducers (4-RSTs), and (B) inspiratory and expiratory volumes, utilizing a soft respiratory mask with a two-way bulb. Speech sounds were recorded for (A). In addition, the subjects' pulmonic vital capacities were measured. Five male students read two kinds of text materials in either Japanese (L1) or English (L2) five times. Thus, we acquired 200 samples for chest and abdominal movements, 100 samples for respiration, and 100 samples for the recorded voice in addition to his pulmonic data and made quantitative and statistical analyses. Our findings were: (1) text reading speed was more stable in L1 than in L2, (2) the lower abdomen was controllable. This may indicate "stomach respiration," (3) the inspiratory and expiratory controls during speech differed largely from those in quiet breathing, indicating acute active muscular movements, (4) the vital capacity volume did not sufficiently correlate with the subjects' expiratory and inspiratory air volume during speech, and (5) a better English pronunciation might be supported by alternative control of inspiration and respiration.

Index Terms: speech production, respiratory muscular control, respiratory control, Japanese L1 and L2 readings, 4-RSTs, a soft-mask with a two-way bulb

1. Introduction

The purpose of this research is to examine (1) how L1 speakers control both their chest and abdominal muscles during inspiration and expiration, and (2) the relationships between pulmonary vital capacity and expiratory and inspiratory air volume, when reading L1 (Japanese) and L2 (English) texts. Japanese is pitch-accented, and English is a stress-timed language. Thus, we hypothesize that stress is realized by both intensity and pitch in English, and as such its subglottal pressure is stronger than Japanese. Additionally, we hypothesize that the so-called "lower abdominal muscular movement" is more active in English than in Japanese. Often, it is claimed in English phonetic education that "abdominal-respiration" is important in its learning, since Japanese speakers (JL1) apt to use "chest-respiration" only. This study will verify how the Japanese learners of English (JL2) differentiate these respiratory manners.

With regard to biological aspects, Williams [1, p. 1651] claims that speech mechanisms are essentially linked to expiratory mechanisms. Human speech requires a source of energy (initiation) by the control of the airstream that is produced by expiratory muscles, mainly by the diaphragm, a structure which oscillates (phonation) and acts as a modulator. However, in speech production expiration does not occur without inspiration.

There are three main dimensions in speech production: respiratory control by muscular control, laryngeal efforts, and articulation. One hypothesis is that all these must be closely connected in a complex manner in speech production. There is a great number of experimental studies on speech production. However, most of them deal with prosodic features, particularly focusing on the fundamental frequency (F0). In our previous studies, we compared the correlation between pitch (F0) with intensity for L1 and L2 speakers and found some differences between them. (E.g., [2-3]) Then we investigated the relationship between acoustic features (pitch, intensity) and respiratory muscular control (of chest and abdomen). To measure the respiratory muscular control, we utilized two thin-wired respiratory strain-gauge transducers (2-RSTs). The reasons for using RST and thin-wired respiratory strain-gauge transducers are described in our previous study in detail (E.g., [4]) We found the same as the previously mentioned relationships, but furthermore, that (1) laryngeal efforts are not highly correlated to respiratory muscular movements (E.g., [5]), (2) different language speakers (L1) use own muscular controls(E.g., [6]), and (3) L2 speakers use different muscular control from L1 speakers. (E.g., [7-8])Saida [9, p. 33] points out that the respiratory muscular movements depend on inspiration and expiration in speech, respectively. Thus, we decided to use a soft respiratory mask with a two-way bulb so that we can measure inspiratory and expiratory air volume respectively during speech in conjunction with RST. Saida [9, pp. 32-33] describes that for stomach respiration, the diaphragm plays the most important role and it moves upward or downward. In addition, abdominal respiratory muscles move together. Thus, we decided to use 4-RSTs, adding two to the above-mentioned previous experiments: one for upper chest and one for lower abdomen, referring to Saida's method [9] using 4-RSTs. Concerning the relationship between respiration/breathing and speech, a number of studies have been published (e.g., [10-25]) focusing on pause (e.g., [10-13]), conversation and or spontaneous speech (e.g., [14-18]), etc., utilizing pneumograph (e.g., [10]), a belt-type RST (e.g., [19]), respiratory (hard) mask, and/or extracted acoustic features, etc. One study [11] gave good insight to our present study comparing the relationship between expiration, chest, and abdominal muscle movements utilizing sensors (presumably, 2-RSTs).

Yet, to our knowledge, there have been no such experiments combining 4-RSTs to measure the upper and lower chest movements, upper and lower abdominal movements, and

 10.21437/Interspeech.2018-1948

a soft respiratory mask with a two-way bulb to measure the inspiratory and expiratory air volume in addition to pulmonic function while reading in mother tongue (L1) and a targeted foreign language (L2).

For this study, our research questions are as follows:

(1) Is there a correlation between the upper and lower chest and abdominal movements?
(2) How are the four respiratory muscular controls related to inspiratory or expiratory control respectively?
(3) How is the inspiratory control related to expiratory control?
(4) Is there any correlate between inspiratory and expiratory volume and L1 speakers' vital capacity volume?
(5) Are there chest respiration and stomach respiration? Can we separate them?

2. Experiments

2.1. Reading Materials

The materials consisted of four kinds: a fable "The North Wind and the Sun" (hereafter NW) in Japanese and English versions adopted from [26], and a fable "Momotaro" ("A peach boy") in Japanese (a concise version) and English (translated version from the Japanese version). In our previous studies we used only NW. Adding another fable was to confirm if "L2 speakers use different muscular control from L1 speakers" is true in this study as well.

2.2. Subjects

The subjects (S1-S5) were five Japanese male students (20-25 years old) who were majoring in English (intermediate and advanced levels).

2.3. Methods

For this study, we conducted three kinds of experiments to acquire data for (A) respiratory function, (B) respiratory muscular controls according to four different materials, and (C) inspiratory and expiratory controls according to four different materials. The data of the five subjects in (A) were collected by a medical expert of respiration. The other two experiments were conducted in the acoustic studio of the first author's University in Japan. In the experiment (B), the subjects' respiratory control and voice were recorded separately. The four thin-wired RST were placed on the subject's upper (Chest1) and lower chest (Chest2), and upper (Abdomen1) and lower abdomen (Abdomen2) to record respective muscle movements during utterance, as shown in Figure 1, together with a headset microphone to simultaneously record a subject's voice. The distance between the mouth and microphone was secured while recording so that intensity would not be affected during the recording (Sampling rate: 44.1 kHz). During speech recordings, the mask was not used since high-quality recording was difficult. In the experiment (C), the same subjects read the same texts wearing four RSTs and holding a soft mask linked to a two-way bulb simultaneously (see Figure 1: right). The subject held the mask by both hands to ensure that air did not leak from the mask. Via the two-way bulb we can acquire inspiratory and expiratory air volume data separately.

All subjects read the text materials five times. Before the recording, the subjects were given enough time to learn to read the material fluently. Thus, in experiment (2) we could acquire 200 samples (5 subjects × 4 kinds of texts (2 in Japanese + 2 in English) × 5 times-repetition × 2 kinds of experiments) for muscular movements and 100 voice data simultaneously. In experiment (3) we could acquire 100 samples (5 subjects × 4 kinds of materials × 5 times-repetition) for respiration. The data samples for analyses totaled 400 in addition to five pulmonic function data. The signals from the chest and abdominal muscle movements and inspiratory and expiratory speech volume (in liters) were directly sent to the PC through LabChart, and the speech signals were sent to LabChart through EDIROL. In our pilot test, we found that posture, either sitting or standing, affected both the respiration and muscular movements both in utterance and non-utterance. (E.g., [27]) We selected the sitting position for the subjects, since it might happen more likely as text readings.

Figure 1: *4-RSTs on the upper and lower chest, and upper and lower abdomen in standing position (left), the same in sitting position (middle) and a soft mask with two-way bulb in sitting position (right).*

2.4. Statistical analysis

Figure 2 shows an example of the signals acquired by LabChart for the muscular movements from the Chest1 (1st row), Chest2 (2nd row), Abdomen1 (3rd row), Abdomen2 (4th row), and speech waveforms (5th row) when a Japanese male speaker sat and read "Momotaro" in Japanese.

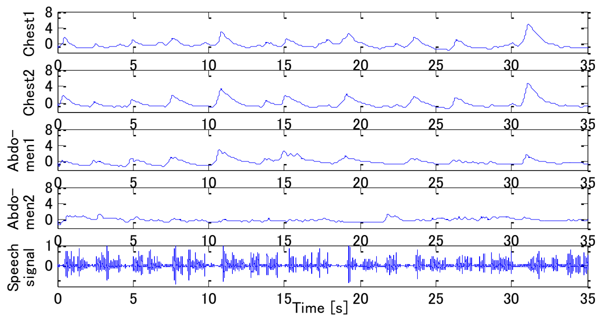

Figure 2: *Signals of 4-RSTs and speech waveform* from *LabChart of a Japanese male speaker's "Momotaro" in Japanese.*

During data processing, we converted all the signals into Matlab data and calculated the mean value of five utterances or overall mean value of all utterances for each muscular movement, inspiratory and expiratory volume, and speech. First, respective signals were down sampled. We subtracted sample means and divided by standard deviation. Next, we used the "detrend" function of Matlab to delete the gap between increased and decreased values. The signals were smoothed by 10-point moving average filters. We extracted each muscular and respiratory peak based on the preset threshold. Thereafter, we calculated the number of peaks per minute of each 4-RST signals and the cross-correlation between two signals at lag 0.

For example, Figure 3 illustrates the muscular movements, displayed by waves, (Chest1, Chest2, Abdomen1, and Abdomen2 from the top), inspiration (5th row) and expiration (6th row) during normal breathing in 40 seconds of a Japanese male speaker (S1) in a sitting position.

Figure 4 illustrates the muscular movements and inspiration, expiration during normal breathing in 40 seconds when S1 was reading a short story "Momotaro" in sitting position. Unlike those seen in Figure 3, the peaks of the abdomens (3rd and 4th) begin steeply at the onset of each sentence or clause. And the peaks of Abdomen1 (3rd) shows the similar manner but they become larger compared to the corresponding signals in Figure 3. The inspiratory peaks become larger or smaller whereas those in Figure 3 are stable.

The numbers of peaks as shown in Figures 3 and 4 are important cues to measure muscular and respiratory control. The red circles in Figures 3 and 4 show the location of respective peak of the line in each row. The peaks of the Chest1 and Chest2 muscles appear to move relatively in the same matter, in accordance with inspiration (5th row). However, the peaks of the abdomen (3rd row) were smaller than those of chests (1st and 2nd). The abdomen 2 shows the least movement, almost none for this speaker's case. The peaks of the chest movements (the 1st and 2nd rows) tend to become larger as time progresses, whereas those of the abdominal movements tended to decline as time progresses.

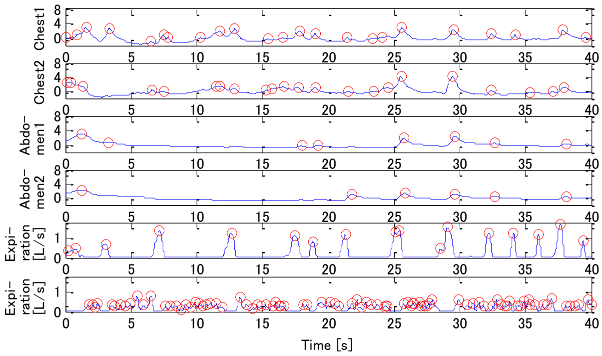

Figure 3: *Muscular muscle movements (upper four lines) and respiratory movement (lower two lines) in 40 seconds during sitting and breathing normally.*

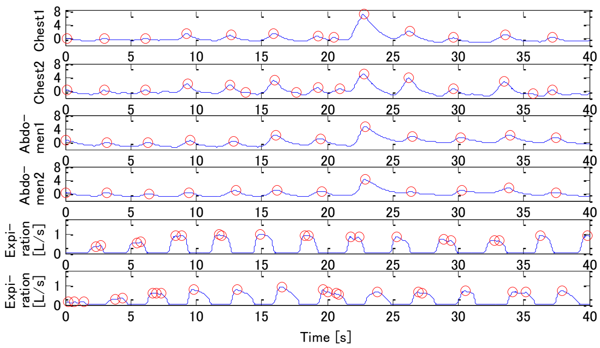

Figure 4: *Muscular muscle movements (upper four lines) and inspiration (5th row), and expiration (6th row) when a Japanese male speaker read "Momotaro" in a sitting position.*

3. Results

3.1. Pulmonic Function

Figure 5 shows each subject's vital capacity volume (= VCV) in liters. S4 had the largest VCV of all subjects. The range among five subjects was 17.5 %. S1 had the largest measured VCV and S5 the least. Their calculated lung age varied from 22 to 64 years. (Cf. 2.2.)

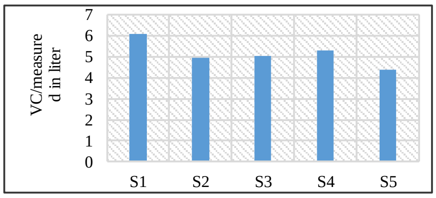

Figure 5: *The subjects' vital capacity volume in liter.*

3.2. Duration of Utterances

Figure 6 exhibits the mean duration (in seconds) of five utterances of each subject on each text. The duration of reading Japanese texts was shorter than that in English. The method to measure the whole duration of each reading time was based on the time when RMS was≧ threshold. The difference in duration may imply each individual's reading speed. Yet, individual difference (variation) in English texts was explicitly larger than in Japanese counterparts. S4 took the shortest time of all.

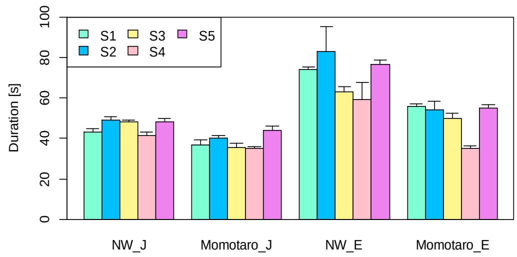

Figure 6: *Mean duration of each subject's five utterances depending on respective L1 and L2 texts.*

3.3. Muscular Control

Figure 7 shows the cross-correlation between Chest1, Chest2, Abdominal1 and Abdominal2 when lag time was 0. Cross-correlation (1) Chest 1 vs. Chest 2 was the highest of all four comparisons. S1 with the largest vital capacity showed the largest correlation in both abdominal movements together with both inspiratory and respiratory.

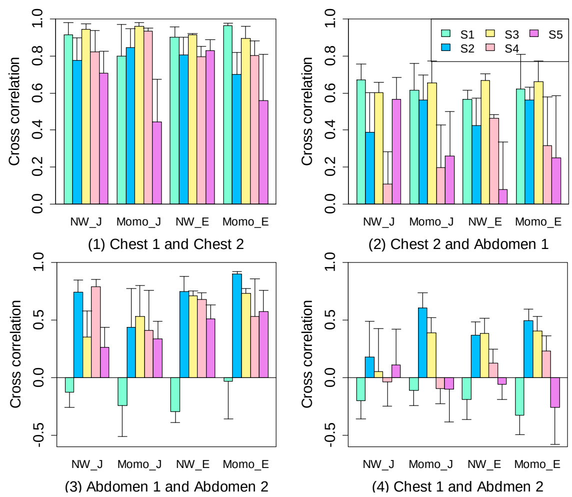

Figure 7: *Cross-correlation between (1) Chest1 vs. Chest2, (2) Chest2 vs. Abdomen1, (3) Abdomen1 vs. Abdomen2, (4) Chest1 vs. Abdomen2, depending on each subject and each text.*

As shown in the Figure, the individual differences appeared more markedly than text differences. Yet, it was clarified that (1) Chest 1and Chest2 moved together, (2) Abdomen1 moved together with the Chest1and Chest2, proving that the diaphragm moves upward and downward and affects the Chest2 and Abdomen1 movements, and (3) Abdomen1 moved relatively independent from Chest1, Chest2 and Abdomen1.

3.4. Respiratory Control

Concerning the relationships between inspiratory and expiratory air volume and vital capacity volume, Figure 8 shows the correlation between the mean totaled inspiratory (left) and expiratory volume (right) and the vital capacity volume (= VCV) for each of the five subjects.

The mean value of the whole expiratory volume (≈ 4 liters) varied more largely than that of the whole inspiratory volume (≈ 2.5 liters). This indicates that the individual difference among the subjects was larger in expiratory volume than inspiratory volume.

In terms of L1 and L2 texts, the inspiratory and expiratory control depended on the subject as indicated in Table 1 based on Figure 8. Table 1 indicates that the subjects used the similar strategies in L1 and L2 texts. In terms of the correlation between the whole inspiratory and expiratory volume and the vital capacity volume (VCV), S2 used the most volume in L2 texts in both inspiratory and expiratory control. S2 used the least volume in L1 text reading in the same control. S2 and S3 did not have higher vital capacity volume than the other subjects. S2's lung age was 64, but it is unclear whether his lung age impacted these results. Thus, our data suggest that the whole inspiratory and expiratory volume used in reading L1 and L2 texts was not necessarily correlated to the vital capacity volume (VCV).

Table 1: *The whole inspiratory and expiratory volume, L1 (J) and L2 (E) texts and the subjects (S1-S5).*

	Whole inspiratory volume (Litter)	Whole expiratory volume (Litter)
S1	J > E	J > E
S2	E > J	E > J
S3	J > E	J > E
S4	J ≈ E	E > J
S5	E > J	E > J

Then, we calculated the correlation between the mean inspiratory and expiratory peak value and the vital capacity volume (= VCV) depending on five subjects as shown in Figure 9. All the subjects used much less peak volume in expiration than in inspiration. All the subjects used much less peak volume in expiration than in inspiration. The mean inspiratory peak value (≈ 0.8 liters) varied more largely than that of the expiratory peak value (≈ 0.4 liters). Particularly, the expiratory peak value was concentrated between 0.25 liters, except for S1's L2 texts (1.2 liter). S1 had the largest VCV of all. S4 showed a remarkable difference from the other subjects between inspiration in L1 (> 2.8 liter) and expiration (0.82 liter) with approximately 2 liters difference.

These data imply that the individual difference among the subjects was larger in the expiratory volume than in the inspiratory volume.S3, S4 and S5 used different respiratory controls depending on L1 and L2 texts (see * on Table 2). Their English pronunciation is much better than that of S1 and S2. It is not certain that their control can verify the subjects' good pronunciation at this stage.

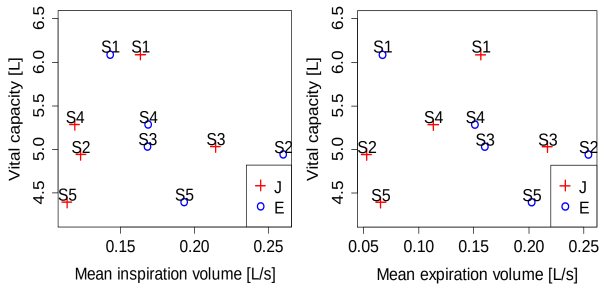

Figure 8: *Correlation between the mean whole inspiratory volume (left) and expiratory volume (right) - y-axis, and the vital capacity volume (= VCV) - x-axis, depending on the subject and on L1 (J) and L2 (E) text.*

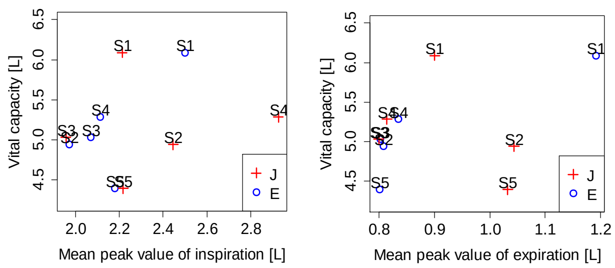

Figure 9: *Correlation between mean inspiratory and expiratory peak value and the vital capacity volume (= VCV) depending on five subjects.*

Table 2: *The peak value in inspiratory and expiratory volume, L1 (J) and L2 (E) texts and the subjects (S1-S5).*

	Inspiratory peak value	Expiratory peak value
S1	E > J	E > J
S2	J > E	J > E
S3*	J > E*	E > J*
S4*	J > E*	E > J*
S5*	E > J*	J > E*

4. Conclusions

From the above experiments, we found that (1) text reading speed was more stable in L1 than in L2, showing a better result than in our previous work by adding another fable, (2) the lower abdomen is controllable as JL1 attempted to move it more actively in L2 text reading than in L2 counterparts . This may indicate "stomach respiration." To reinforce this indication, we need to examine English speakers. (3) the large inspiratory and expiratory differences between quiet breathing and in speech indicate acute active muscular movements in speech, (4) the vital capacity volume did not sufficiently correlate the subjects' expiratory and inspiratory air volume during speech, despite the fact that the subjects were young, but their lung age varied largely, (5) a better English pronunciation might be supported by alternative control of inspiration and respiration.

For further studies, we need to use much longer passages than those in the L1 and L2 texts used for this study and reinforce these results.

5. Acknowledgements

This study is partly supported by Grant-in-Aid KAKENHI for Scientific Research C (NO. 17K02698) of Japan.

6. References

[1] P. L. Williams, *Gray's Anatomy: The Anatomical Basis of Medicine and Surgery,* 38e *(Gray's Anatomy: the Anatomical Basis of Clinical Practice)*, Churchhill Livingstone, 1995.

[2] T. Isei-Jaakkola, "Cognition and Physio-acoustic Correlate: Audio and Audio-visual Effects of a Short English Emotional Statement: on JL2, FL2 and EL1, " *Advances in Natural Language Processing, Lecture Notes in Computer Science*, 4139 2006, pp. 161-173, Springer, 2006.

[3] T. Isei-Jaakkola, Q. Sun, and K. Hirose, "Audio and audio-visual effects of a short English emotional sentence on Japanese L2s' and English L1s' cognition, and physio-acoustic correlate, " *Proceedings of International Conference of Speech Prosody Dresden*, pp. 225-228, 05.2006.

[4] T. Isei-Jaakkola, Y. Nagano-Madsen, and K. Ochi, "Respiratory Control, Pauses, and Tonal Control in L1's and L2's Text Reading – A Pilot Study on Swedish and Japanese –," *Proceedings of International Conference of Speech Prosody 2018,* pp. 873-877, 06.2018.

[5] T. Isei-Jaakkola and K. Hirose, "The relationships between, pitch, intensity, vocal fold vibrations and chest- and stomach-muscle movements in a read speech (1)," (in Japanese) *Proceedings of Autumn Meeting,* Acoustical Society of Japan, pp. 269-270, 2013.

[6] T. Isei-Jaakkola and G. Ohyama, "Respiratory pressure control and linguistic rhythm utilizing Respiratory Strain-gauge Transducer—A pilot test on JEFL and EL1—," (in Japanese) *Bulletin of Humanities,* Chubu University No. 23, pp. 81-89, 03.2010.

[7] T. Isei-Jaakkola, "A linguistic rhythm observed from the respiratory muscle movements and speech waveforms by English, Japanese and Chinese L1 and L2 – the recitation of a story," *Proceedings of the 17th ICPhS* Hong Kong, pp. 942-945, 2011.

[8] T. Isei-Jaakkola and K. Hirose, "A cross-linguistic comparison of the chest and stomach muscle movements of JL1, JL2, PL1, and PL2 while reading a short story," *EAA proceedings*, Forum Acusticum 2014 (7th Forum Acusticum Krakow 2014), Krakow, Poland, 6 pages in CD (ISSN: 2221-3767).09.2014.

[9] H. Saida, *Medical Voice Designer Assists You In Your Vocalization Using Newly Developed Voice Maps* (in Japanese), Ongaku No Tomo Sha, 2016.

[10] B. Conrad and P. Schönle, *Speech and respiration*, *Archiv für Psychiatrie und Nervenkrankheiten*, *226*(4), pp. 251-268. (1979)

[11] M. Sugito, G., Ohyama, and H. Hirose, "A preliminary study on pauses and breaths in reading speech materials," *Annual Bulletin RILP*, *24*, pp.121-130, 1990.

[12] F. Grosjean and M. Collins, "Breathing, pausing and reading," *Phonetica*, *36*(2), pp. 98-114, 1979.

[13] J. Trouvain, C. Fauth, and B. Möbius, "Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech," *Proceedings of Speech Prosody (SP8)*, *31*, pp. 31-35, 2016.

[14] M. Denny, "Periodic variation in inspiratory volume characterizes speech as well as quiet breathing," *Journal of Voice*, *14*(1), pp. 34-46, 2000.

[15] F. Torreira, S. Bögels, and S. C. Levinson, "Breathing and speech planning in turn-taking," In *the 10th International Seminar on Speech Production: Satellite Workshop on "Interpersonal coordination and phonetic convergence,"* 2014.

[16] S. Fuchs, C. Petrone, A. Rochet-Capellan, U.D. Reichel, and L.L. Koening, "Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands," *Journal of Phonetics 5*2, pp. 35-45, 2015.

[17] A. Rochet-Capellan and S. Fuchs, "Take a breath and take the turn: how breathing meets turns in spontaneous dialogue," *Phil. Trans. R. Soc. B*, *369*(1658), 2013.

[18] A. Mendes, L. Ferreira, A. Jorge, H. Rothman, D. Tavares, and I. Garcez, "Respiratory physiology during reading aloud tasks," *Distúrbios da Comunicação*, *29*(3), pp.519-528, 2017.

[19] C. Dromey and L. O. Ramig, "Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation," *Journal of Speech, Language, and Hearing Research*, *41*(5), pp.1003-1018, 1998.

[20] D. H. Whalen and J. M. Kinsella-Shaw, "Exploring the relationship of inspiration duration to utterance duration," *Phonetica*, *54*(3-4), pp.138-152, 1997.

[21] M. Włodarczak and M. Heldner, "Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking," *Proceedings of INTERSPEECH 16*, San Francisco, USA, pp. 510-514, 2016.

[22] M. Włodarczak, M. Heldner, and J. Edlund, "Communicative needs and respiratory constraints," *Proceedings* of *INTERSPEECH 2015, Dresden, Germany,* pp. 3051-3055, 2015.

[23] D. H. McFarland, "Respiratory markers of conversational interaction," *Journal of Speech, Language, and Hearing Research*, *44*(1), pp.128-143, 2001.

[24] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, "Acoustic and respiratory evidence for utterance planning in German," *Journal of Phontics*, *41*(1), pp.29-47, 2013.

[25] M. Włodarczak and M. Heldner, "Respiratory Constraints in Verbal and Non-verbal Communication," *Frontiers in psychology*, *8*, p. 708, 2017.

[26] T. Isei-Jaakkola, K. Hirose, and A. Itai, *Proceedings of Spring Meeting,* Acoustical Society of Japan, pp. 1383-1386, 03.2018.

[27] International Phonetic Association, *Handbook of the IPA*, Cambridge University Press, 1999.