A PRELIMINARY ACCOUNT OF INBREATHS IN ROMANIAN SPONTANEOUS SPEECH*

OANA NICULESCU[1]

Abstract. Human speech is a blend between verbal and non-verbal vocalisations (NVVs). Based on previous studies pertaining to different research fields, this article focuses on breath intakes occurring in spontaneous speech. Classified as NVVs, alongside other phenomena such as vegetative sounds, affect bursts, onomatopoeia, filler particles and melodic utterances, inbreaths perform multiple and often cumulative roles in the discourse. Moreover, the interplay between speech planning and breathing control has been extensively researched in numerous studies. Analyses of NVVs in human speech have the potential of bridging the gap between various research domains, ranging from psycholinguistics and clinical linguistics to discourse management and speech planning, forensic voice comparison, as well as language processing. While the topic remains relatively unexplored in Romanian linguistics, our current study aims to broaden our understanding of the distribution and acoustic correlates of inbreaths in monologue speech.

Keywords: Romanian connected speech, breath intakes, non-verbal vocalisations.

1. INTRODUCTION

The aim of this study is to draw attention to a relatively unexplored research topic in Romanian linguistics related to breath intakes in spontaneous speech. Based on previous studies pertaining to various research fields, our study first focuses on inbreaths as an integral part of breathing patterns (Section 2.1). An overview on respiratory rates computed from (un)healthy subjects is provided, with images derived from Yuan *et al.* (2013) illustrating the differences between normal tidal breathing compared to abnormal breathing patterns. We then explore how respiratory kinematics in speech differ from those employed in tidal breathing (Section 2.2). More precisely, we investigate breath intakes occurring in spontaneous speech showcasing the main results obtained on data derived from well-resourced Germanic languages.

---

* Research Grant GAR2023 (168), DIS-Ro project, contract number 758/23.11.2023, with the financial support of the Donor Recurrent Fund, at the disposal of the Romanian Academy and managed through the "PATRIMONIU" Foundation.

[1] Romanian Academy Institute of Linguistics "Iorgu Iordan – Alexandru Rosetti", oeniculescu@yahoo.com.

In what follows, we focus on the way in which inbreaths are depicted in Romanian conversational data, advocating for the inclusion of non-verbal vocalisations in the annotation system (Section 3).

Our contribution resides in providing acoustic illustrations of breath intakes in Romanian monologue speech (Section 4). Audible inbreaths are depicted in a wide range of pausal contexts, either patterning with silent phases and tongue clicks, or surfacing in the offset of a laughter episode. Previously unattested cases of multiple inbreaths occurring in one inspiratory breathing episode are also discussed. The article concludes with an outlook and future research paths (Section 5).

2. PREVIOUS RESEARCH ON INBREATHS IN HUMAN SPEECH

2.1. Respiratory rates

We first investigate inbreaths from a physiological perspective, as an integral part of breathing patterns. Differences in respiratory rates are conditioned by normal and abnormal breathing patterns. Figure 1 illustrates these patterns as depicted by Yuan *et al.* (2013).

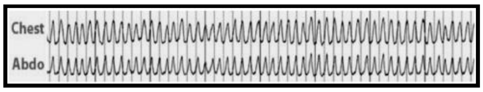
Normal tidal breathing, also referred to as diaphragmatic breathing, is comprised of upward deflection (i.e., inspiration) and downward deflection phases (i.e., expiration) reflected by the synchronized movement of the thorax and abdomen (Figure 1a), while asynchronous movements (thoracic or paradoxical) are associated with abnormal breathing patterns (Figure 1b–h), indicative of various breathing pattern disorders. Normal respiratory rates range from 12 breaths per minute to 26 breaths per minute in adults (Chon *et al.* 2009), while these values vary in children according to age, with a declining rate from birth to early adolescence (Fleming *et al.* 2011).

Studies calculating respiratory rates mainly rely on ECG or PGG signals[2]. Based on a time-frequency spectral estimation method derived from photoplethysmogram signals used as a commonly non-invasive practice for measuring oxygen saturation in the blood, Clifton *et al.* (2007) computed and compared respiratory rates from both healthy and unhealthy subjects. Their results varied from 6.2 to 35.8 breaths per minute (bpm).

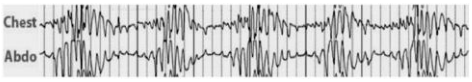
Results pertaining to the three healthy volunteers revealed significantly lower respiratory rates, ranging from 6.2 to 12.7 bpm, compared to those obtained from patients suffering from various respiratory conditions (11.1 to 35.8 bpm).

Patients who suffer from type II respiratory failure had the highest rates (25.0 to 35.8 bpm), followed by subjects who presented pneumonia (20.5 bpm) or acute asthma (16.3 bpm). Patients under general anaesthetic display an average 12.4 bpm ratio, while those recovering from general anaesthesia vary from 13.7 to 26.3 bpm. Typically, most (un)healthy subjects have a spontaneous breathing rate falling within 0.2 and 0.4 Hz, while a respiratory frequency above 0.5 Hz is extremely rare (Chon *et al.* 2009).

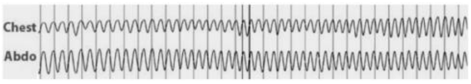
[2] For an in-depth review of the literature on breathing rate estimation derived from the electrocardiogram (ECG) and pulse oximetry signals, i.e., photoplethysmogram (PPG), see Charlton *et al.* (2017).



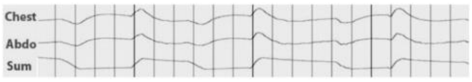
(a) *Normal tidal breathing*
symmetric chest wall and abdomen movement



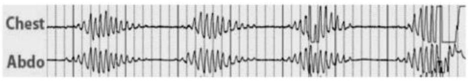
(b) *Thoracoabdominal paradox*
asynchronous thorax and abdomen movement



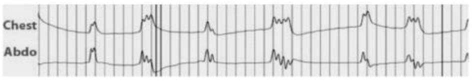
(c) *Simulated Kussmaul's breathing*
increased frequency and tidal volume indicative of a rapid and large amplitude



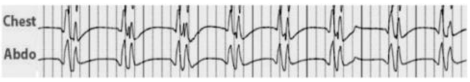
(d) *Simulated apneustic breathing*
inspiration and expiration are each followed by prolonged pauses



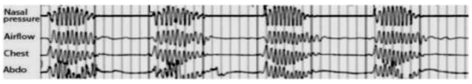
(e) *Cheyne-Stokes respiration*
cyclical crescendo-descendo pattern of chest wall and abdomen movement



(f) *Simulated ataxic breathing*
irregular chest and abdominal movements with unpredictable pauses between inspiration and expiration



(g) *Biot's breathing*
regular large amplitude of chest and abdominal movement scattered across episodes of apnea



(h) *Central apnea*
breathing which occurs during sleep characterized by the lack of nasal flow and pressure correlated with an absent chest and abdominal movement effort

Figure 1. *Normal and abnormal breathing patterns*
Images extracted from Yuan et al. (2013) with each large box representing 30 seconds

2.2. Inbreaths in speech

Respiratory kinematics in speech differ from those employed in tidal breathing. When investigating the temporal patterns of inbreaths and outbreaths in 3 individual conditions pertaining to quiet breathing, listening within a conversational framework, and speech (reading aloud, spontaneous conversation), McFarland (2001) demonstrated that speech is associated with more rapid breath intakes and a prolonged expiratory phase. Furthermore, the study showed that there is a significant and consistent reduction in inspiratory duration during speech as compared to quiet breathing. In turn, differences in expiratory duration were highly speaker-dependent, with outbreaths displaying a tendency of increasing from quiet breathing to reading aloud and further to spontaneous monologues. During listening, inspiratory duration decreased, resembling speech values more closely. Another important result stems from the observation that, to a certain degree, respiratory patterns are shaped by communicative demands. Moreover, during turn-taking events, participants displayed a synchrony between the respiratory kinematics.

Apart from the temporal patterns associated with breathing kinematics in various conditions, there are also differences in terms of airflow through the nasal and oral pathways. If tidal breathing is characterised by inspiratory nasal airflow (lips adducted, mouth closed, negative nasal pressure), breath intakes in speech have been observed to follow different patterns. Opting for a nasal cannula in favour of a face mask, Lester and Hoit (2014) determined that the typical pattern for breath intake in healthy adults was simultaneous nasal and oral inspirations. These findings were consistent across all speaking tasks including counting (from 1 to 50), paragraph reading (California Passage), spontaneous speaking (one minute), and conversation (phone conversation with a friend or family member during the session). It is also worth noting that participants were seemingly unaware of this trend and did not consider it to be their natural breathing pattern.

Breath noises in human speech have been classified as non-verbal vocalisations (NVVs) alongside other phenomena such as vegetative sounds (swallowing, coughing and throat clearing, snoring, sneezing, etc.), affect sounds (laughing, crying, screaming) or affect bursts in a broader sense, defined as "short emotional non-speech expressions interrupting speech" (Schröder 2003), interjections and filler particles (Belz 2023), melodic utterances (whistled, hummed or sung, cf. Trouvain and Truong 2012). Among NVVs, audible breath intakes are considered to be the most frequent (Trouvain and Truong 2012; Trouvain *et al.* 2016, among others).

When investigating the acoustic correlates of inbreaths in read speech compared to data derived from spontaneous speech, Trouvain *et al.* (2019) observed that breath intakes in read speech display COG values of under 2 kHz as well as a low intensity. In spontaneous speech, inbreaths have been observed to appear in turn initial positions, sometimes patterning with tongue clicks. Due to the exploratory nature of the study, certain aspects of breath intakes, such as distribution and temporal patterns according to speaking style, were not fully explored.

Studies have also highlighted the interplay between speech planning and breathing control. Based on data collected from 26 female speakers recorded using an Inductance Plethysmograph, Rochet-Capellan and Fuchs (2013) have shown that the structure of the breath group (i.e., interval of speech produced on a single exhalation) in German spontaneous speech is reflected in breathing kinematics. More precisely, their findings reveal that: (1) inhalations surface at syntactic boundaries, either before a matrix or an embedded clause, (2) the duration of a breath intake (averaging at 0.7 s) and the amplitude (17.6%MD) depended on the length of the breathing group (averaging at 3.5 s) as well as on the type of the first clause (53% matrix clause, 24% embedded clause, 23% incomplete clause), (3) inhalation was significantly deeper before a matrix clause (+3.5MD) and also before a breath group with at least one hesitation (especially in the onset), (4) inbreaths were deeper and longer prior to a disfluency (hesitation, repetition, repair, uncomplete clause). Furthermore, similar amplitudes of exhalations and inhalations were found in breath groups of 15-18 syllables (or with 2 clauses), considered to be the preferred patterning between linguistic structure and breathing kinematics in spontaneous speech.

The link between kinematic behaviours, as for example hand movements, facial expressions, gaze, gestures, and the acoustic correlates of non-verbal vocalisations has not yet been fully explored. Most studies in this line of research tend to focus on the blend between kinetic activities and NVVs such as clicks and inbreaths based on interactional data, especially at turn-taking or turn-management phases (Schegloff 1996; Wright 2011; Ogden 2013; Torreira *et al.* 2016; Pinto and Vigil 2019; Kosmala 2020). The main findings suggest that kinetic events associate differently with clicks and inbreaths, supporting various functions such as word searches, discourse management or speech planning.

From a pragmatic perspective, inbreaths, alongside other NVVs, have been the focus of studies dealing with turns grounded in discourse analysis with results indicating that respiratory noises, surfacing at specific prosodic and syntactic boundaries, can function as discourse markers signalling a turn-taking or a turn-yielding cue (McFarland 2001; Schegloff 2006; Fuchs *et al.* 2013; Rochet-Capellan and Fuchs 2013; Torreira *et al.* 2016, among others). Furthermore, in their socio-phonetic study related to the properties of formal and informal speech registers in Korean, known for its honorification system (Sohn 1999; Yoon 2004; Byon 2006), Winter and Grawunder (2012) observed that acoustic correlates of politeness (associated with formality) patterned with inbreaths. More precisely, their results showcase that noisy breath intakes, having an ingressive "hiss" of up to 20 kHz, were most prevalent in the formal condition, particularly among male speakers, who registered 4.1 times more "hisses" in the formal compared to the informal condition. The authors concluded that breath intakes performed a dual function related to the physiological demand for air intake, as well as conveying "social meaning" by calibrating to the social context of an utterance.

Studies in forensic phonetics focusing on breathing patterns in speech are still rare. Among the handful of research on this topic, Kienast and Glitza (2003) showed that non-pathological respiratory sounds such as inbreaths and outbreaths display speaker-specific differences. These differences were observed in regard to the total number of respiratory sounds, as well as in connection to the spectral properties and various patterns of inspiratory or expiratory airflow.

Initially discarded from ASR systems, NVVs in general, and especially breath noises and (speech) laughter are now being acknowledged and integrated into newer automatic speech recognition systems, paving the way to more natural-like dialogue systems, as well as broadening the scope of various speech applications. Breath intakes in speech pauses have also received attention from speech synthesis research, where modelling, for instance, pauses in audiobooks synthesis can determine an increase in speech naturalness (cf. Braunschweiler and Chen 2013).

This overview of work on inbreaths pertaining to different research discipline has showcased that breath intakes, alongside other non-verbal vocalisations, perform multiple and often cumulative roles in human speech. Our focus now shifts to the way in which inbreaths are depicted in Romanian conversational data.

3. ROMANIAN SPEECH DATA

Romanian conversational speech corpora, either native (CORV, IVLRA, IV II CORTES, ROVA, CLRVA I–II, CLRV, among others) or non-native (Constantinescu and Stoica 2020) have not yet employed systematic annotation schemas to incorporate breath intakes along with other verbal and non-verbal vocalisations. Even though the transcriptions proposed by Sacks, Schegloff and Jefferson (1974) and further perfected by Jefferson (1978,

2004) include both inbreath and outbreath episodes[3], this annotation was not assumed when transcribing Romanian conversational data[4]. The same observation holds true for dialectal data, with most of the texts derived from the National Phonogram Archive of the Romanian Language (Şuteu1958). We argue that this methodological preference is motivated, to a certain extent, by the low-quality audio recordings, especially those pertaining to dialectal field inquiries, which can hinder analyses on NVVs such as inhalations and exhalations. Additionally, as demonstrated by Belz and Trouvain (2019), what where previously considered (and transcribed) as silent pauses, are, in fact, pauses which predominantly contain inbreaths or tongue clicks. Thus, it is highly probable that the annotations used in Romanian speech corpora for pauses actually contain various non-verbal vocalisations, in addition to completely silent pauses.

Preliminary findings (Table 1) stemming from a broader examination of annotation practices in Romanian speech corpora, similar to the one undertaken for conversational English by Trouvain and Truong (2012), reveal that: (1) transcriptions do not include a distinct category for breath noises, (2) when transcribed, inbreaths are marked between doubled parentheses, containing the transcriber's observations and descriptions, thus being placed alongside other verbal and non-verbal vocalisations, face and body gestures or even extra-linguistic information related to recording conditions, (3) breath intakes have an extremely low frequency of occurrence compared to the overall duration of the recorded material, with most of the corpora registering only one or two occurrences. IVLRA has the largest number of annotated inbreaths (N = 40), while CORV, CORTES and even the recently developed multimodal CLRV corpus (in print version only) do not incorporate any cases of inspiratory or expiratory episodes.

Corpus	Duration/ Texts	N
CORV (2002)	220 min/37	0
IVLRA (2002)	35 texts	40
CORTES (2009)	24 texts	0
Gheorghe *et al.* (2009)	334 min/ 29	1
ROVA (2011)	33 texts	1
CLRVA II (2013)	78 texts	2
Constantinescu and Stoica (2020)	79 oral texts	1
CLRV (2023)	97 texts	0

Table 1.
Distribution of breath intakes in Romanian conversational corpora

Benefiting from high-quality audio recordings and time-aligned speech based on TextGrids in Praat, newer Romanian speech corpora, such as the Ro-Phon corpus (Niculescu 2021, 2022) developed through the ROC-lingv project (2020–2022, UEFISCDI funding) and

[3] The dot-prefixed 'h's indicate an inbreath (.hhh), while an outbreath is annotated by 'h's without the dot (Jefferson 2004: 40).

[4] It should be noted that the Jeffersonian transcription system, highly regarded and widely used for transcribing Romanian speech data, was designed with a focus on conversation analysis and therefore inherently reflects the fundamental theoretical principles of this linguistic framework. Furthermore, like any annotation schema, it is "necessarily selective" (Atkinson and Heritage 1984).

the currently in development corpus pertaining to the DIS-Ro project (2024–2025, Romanian Academy funding), have integrated NVVs in their transcription system. Even though breath noises were not initially a distinct category in the Ro-Phon corpus annotation schema, the DIS-Ro project extended the annotation to include all vegetative sounds. By including non-verbal vocalisations in the transcription system of Romanian conversational corpora we can gain a better understanding of the mechanisms employed in discourse management and speech planning.

In what follows, we present acoustic illustrations of inbreaths occurring in Romanian connected speech based on data pertaining to the re-annotated Ro-Phon corpus. It should be noted that the transcription system is still being development and can undergo further modifications in the future.

4. ACOUSTIC ILLUSTRATIONS OF BREATH INTAKES IN ROMANIAN CONNECTED SPEECH

To the best of our knowledge, there are no documented cases of breath intakes pertaining to Romanian speech data described from an acoustic perspective. Consequently, in this section, we provide a preliminary phonetic analysis of inbreaths in monologue speech. In order to make the data accessible, we include QR codes for each example, linking the illustrations with the corresponding audio files.

Serving as a measure for comparison, Figure 2 depicts a completely silent pause in a phonetic sense, which does not contain any phonetic particles or events (cf. Belz and Trouvain 2019). With a duration of 497 ms, this pause would fall within the medium-length range according to the classification put forward by Campione and Véronis (2002)[5].

Most inbreaths occurring in fluent speech are audible, and thus visible on the spectrogram. Nevertheless, there are cases in which breath intakes are less audible, being produced with a lax vocal tract. An instance such as this is presented in Figure 3, the breath intake extending to 397 ms, with a low intensity of 29 dB, and COG values reaching 418 Hz. Classified by Mettouchi (2018) as reflex inbreathing (in opposition to voluntary or controlled audible breath intake, i.e, audible inbreaths), this type of inhalation noise is usually discarded from large-scale corpora analyses (Kienast and Glitza 2003; Fuchs *et al.* 2013; Rochet-Capellan and Fuchs 2013; Trouvain 2014; Torreira *et al.* 2016; Trouvain *et al.* 2016, 2019; Kosmala 2020, among others).

As observed in previous research, breath intakes typically surface between two stretches of fluent speech delimited by brief silences on either side. This case is illustrated in Figure 4, each silence lasting less than 60 ms, with the left edge silence being shorter in comparison to the one placed on the right edge of the inbreath. In this example, it appears that the duration of the audible inbreath is similar to that of the inaudible breath intake. However, there is a difference in intensity and COG values between the two outputs, the strong inbreath having higher values, of 44 dB and 2737 Hz, respectively.

[5] Measuring approximately 6k pauses in read and spontaneous speech from various spoken corpora, Campione and Véronis (2002) observed a trimodal distribution in terms of pausal duration, distinguishing between short (<200 ms), medium (200 – 1000 ms) and long silent pauses (>1000 ms). Pauses longer than one second surfaced only in spontaneous speech.

Our data reveal that breath intakes in monologue speech pattern with silent pauses of various length, thus corroborating Mettouchi's (2018) findings. In line with this observation, in our speech corpus, we identified inhalation noises followed (Figure 5) or preceded by silence (Figure 6), as well as surfacing between two stretches of silence (Figure 7). There were also cases, though with a lesser frequency of occurrence, where inbreaths surfaced directly attached to speech on either side (Figure 8), without the short silent edges. The breath intake in this particular context has the shortest duration (319 ms) among all inbreaths presented so far.

Another case attested in the literature is when inbreaths pattern with non-phonemic clicks[6], a context also identified in our corpus. Figures 9 and 10 illustrate breath intakes with a predominantly nasal airflow intake in the onset followed by a short click, marking the transition toward the oral airflow intake in the offset. The inbreath showcased in Figure 9 has an overall duration of 988 ms, with the nasal section extending to 371 ms, a 30 ms click followed by the oral section of 587 ms. COG values are much higher in the onset of the inbreath (4298 Hz) compared to the offset (859 Hz). In terms of amplitude, the nasal portion registered 45 dB, the click rose to 54 dB, while a decrease to 47 dB was observed in the oral airflow intake. Similar patterns of COG and intensity were observed in the second example (Figure 10), the only variation being a longer duration of the nasal onset (398 ms) compared to the oral offset (292 ms). This type of inbreath displays a sequential nasal – oral airflow intake setting it apart from previously outlined breath intakes characterised by simultaneous nasal and oral inspirations.

Studies on breath intakes and other non-verbal vocalisations occurring in spontaneous speech have described and measured both audible and inaudible inbreaths, with or without clicks in the initial, medial or final portion of the inhalation. However, to our knowledge, there have been no documented cases thus far of multiple breaths intakes occurring in a single inspiratory breathing episode. Inbreaths of this specific type are depicted in Figures 11 and 12. The first pattern identified (Figure 11) consisted of an oral breath intake in the onset, a nasal inhalation at the core, followed by an oral airflow intake in the offset. The overall duration of the complex inhalation was 903 ms. All three portions are of similar length, with the nasal section displaying the highest amplitude and COG values (Table 2a). The second pattern observed in our data (Figure 12) had a primary nasal inhalation in the onset, an oral breath intake at the core, and a secondary nasal inhalation in the offset. The inbreath lasted for 1.1 seconds, followed by a silent pause of 369 ms. Temporal and frequency patterns (Table 2b) were similar to those observed in the previous example, with the onset nasal airflow intake having the highest values for all measurements.

(a) oral – nasal – oral inbreath

	ORAL (onset)	NASAL (core)	ORAL (offset)
Duration (ms)	263	293	348
Intensity (dB)	30	40	30
COG (Hz)	1609	1676	1411

(b) nasal – oral – nasal inbreath

	NASAL (onset)	ORAL (core)	NASAL (offset)
Duration (ms)	376	354	341
Intensity (dB)	35	32	26
COG (Hz)	4454	2595	793

Table 2. *Acoustic measurements of complex breath intakes*

[6] In our data, clicks were found to pattern also with swallowing episodes, especially in the offset (observation in line with Ogden 2021).

Further investigation is required to ascertain the rate at which these complex inbreaths occur in Romanian connected speech. Moreover, dealing with these complex inbreaths also presents a challenge for the annotation system, as it prompts the question of whether to transcribe them as one single inspiratory episode or as separate breath intakes. As we further develop the transcription system within the DIS-Ro project, we will address this question and strive for a suitable answer.

Surfacing in the editing phase of an identical repetition[7], e.g. [je je] 'it is it is' (Figure 13), the inbreath, lasting for 512 ms, is preceded by an outbreath of 551 ms and followed by a silent pause of 150 ms. Demarcating the two respiratory episodes is a tongue click, with a duration of 124 ms. We observe that the amplitude of the exhalation (62 dB) is much higher compared to that of the inhalation (40 dB).

Expiratory and inspiratory episodes can also surface in laughter. More precisely, outbreaths can occur at the onset of a bout, while a deep inbreath often represents the offset of a laughter episode (cf. Bachorowski *et al.* 2001; Trouvain 2003, 2014; Trouvain *et al.* 2019). Figure 14 showcases an entire standalone voiceless laugh[8] which lasts for 2.3 seconds. In this context, the offset inhalation extends to 476 ms, which is almost twice the length of the exhalation onset at 211 ms[9]. From a distributional perspective, our preliminary findings indicate that inbreaths occurring in laughter episodes have a lower frequency of distribution compared to breath intakes surfacing in pausal contexts (with or without tongue clicks).

All of the above examples have mainly depicted inbreaths as standalone episodes. However, spontaneous speech is a blend between verbal and non-verbal vocalisations. With a window length of 3.3 seconds, Figures 15 and 16 illustrate the interaction between audible inbreaths and other NVVs in connected speech.

The extended speech break showcased in Figure 15 encapsulates a deep inbreath of 755 ms followed by a medium-length silent pause and an outbreath lasting for 351 ms. The outbreath was further succeeded by an inter-pausal click. An even broader cluster of NVVs is outlined in Figure 16 containing breath intakes, tongue clicks, silent pauses and other vegetative sounds equivalent to coughing and swallowing. This stands to prove that non-verbal vocalisations tend to pattern together forming extending clusters similar to the way in which "fluencemes"[10] do.

---

[7] For an in-depth study on immediate and identical self-repeats in Romanian monologue speech, see Niculescu and Candea (2023).

[8] Since there is no periodical structure present in the waveform, the laughter episode is classified as unvoiced. Studies have shown that laughter episodes in conversational data tend to be voiceless (cf. Bachorowski *et al.* 2001; Trouvain 2003, 2014; Trouvain *et al.* 2019).

[9] Truong *et al.* (2019) proposed a six-level annotation schema for complex laughter in speech, where inbreaths ("INBR") are transcribed on the respiratory level (1st level, "SPEECH-RESP"), alongside outbreath noises ("OUTBR") and speech ("SP").

[10] "A fluenceme is an abstract and idealized feature of speech that contributes to the production or perception of fluency, whatever its concrete realization may be. Fluencemes build the basis for the distinction between productive, perceptive and nonverbal fluen cy" (Götz 2013: 8).

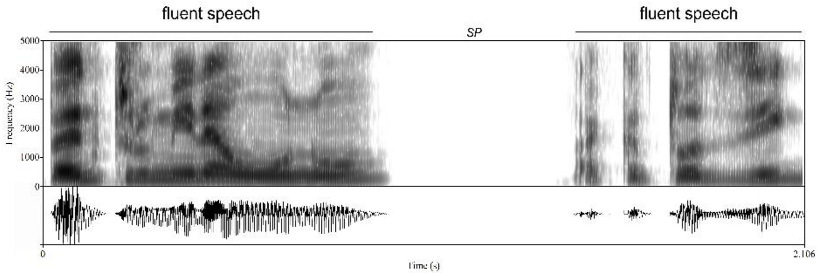


Figure 2. *Silent pause between two stretches of fluent speech*
Broadband spectrogram and waveform of the utterance
pentru mine unu(l) # (0.497) dacă tot zic
"for me <silent pause> if I say"

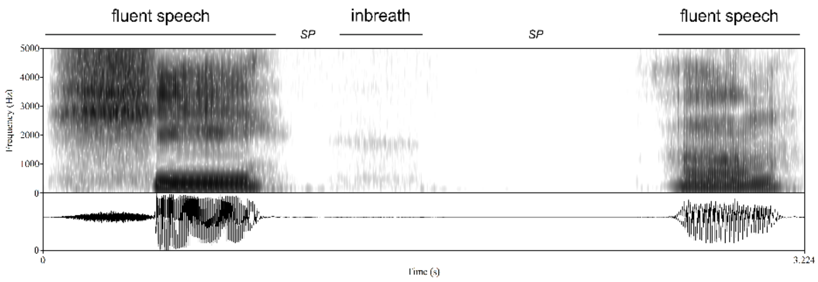


Figure 3. *Inaudible breath intake*
Broadband spectrogram and waveform of the utterance
ş:i: # (0.156) |⇓| (0.397) # (1.038) @ăm
"and <silent pause> <inaudible inbreath> <silent pause> uhm"

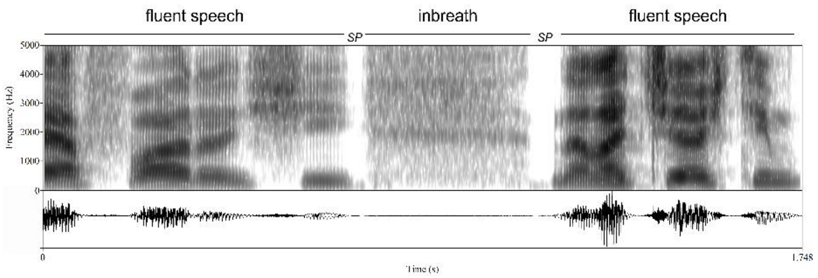


Figure 4. *Audible breath intake between two stretches of fluent speech*
Broadband spectrogram and waveform of the utterance
de afară şi⊥ # (0.036) ⇓ (0.382) # (0.055) da(r)-(î)n acelaşi timp
"from outside and <abandoned phrase> <silent pause> <audible inbreath>
<silent pause> but at the same time"

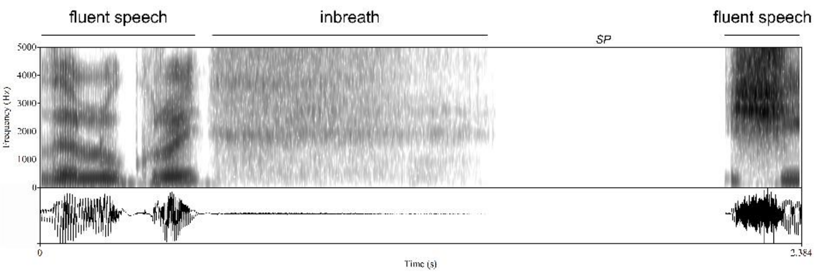


Figure 5. *Audible inbreath followed by silent pause*
Broadband spectrogram and waveform of the utterance
de lucruri ⇓ (0.903) # (0.716)
şi "of things <audible inbreath> <silent pause> and"

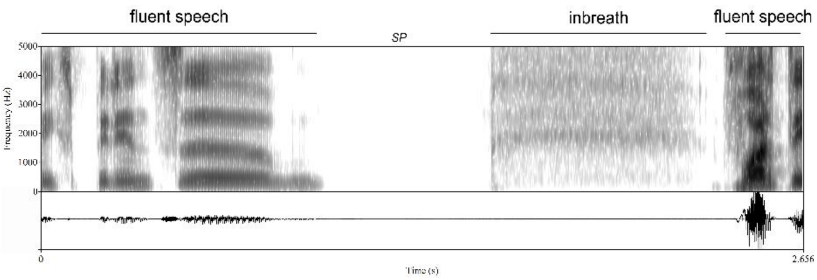


Figure 6. *Audible inbreath preceded by a silent pause*
Broadband spectrogram and waveform of the utterance
experienţă‿@m # (0.582) ⇓ (0.766) foarte
"an experience um <silent pause> <audible inbreath> very"

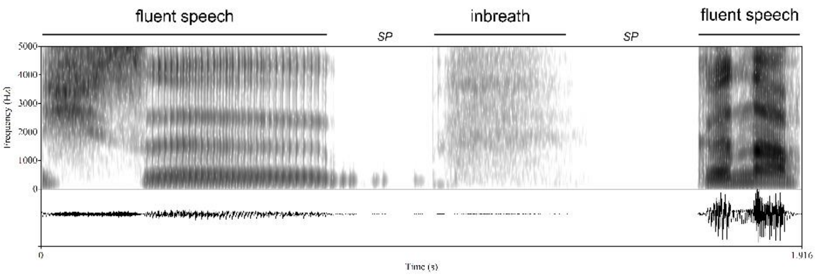


Figure 7. *Audible inbreath surfacing between two silent pauses*
Broadband spectrogram and waveform of the utterance
ş(i) să:⊥ # (0.242) ⇓ (0.406) # (0.261) că &na
"and to <abandoned phrase> <silent pause>
<audible inbreath> <silent pause> because"

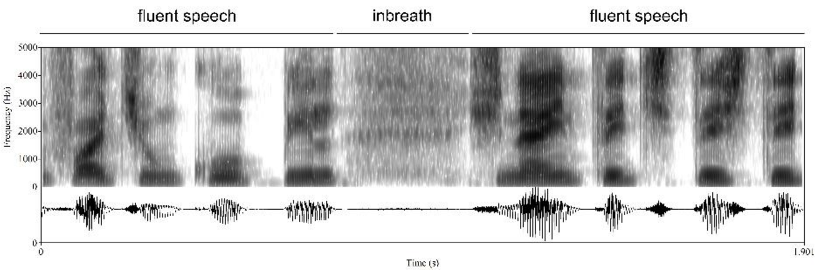


Figure 8. *Breath intake occurring in fluent speech*
Broadband spectrogram and waveform of the utterance
face-un copil ⇓ (0.319) şi-(î)nainte citeşte
“bears a child <audible inbreath> and before reads”

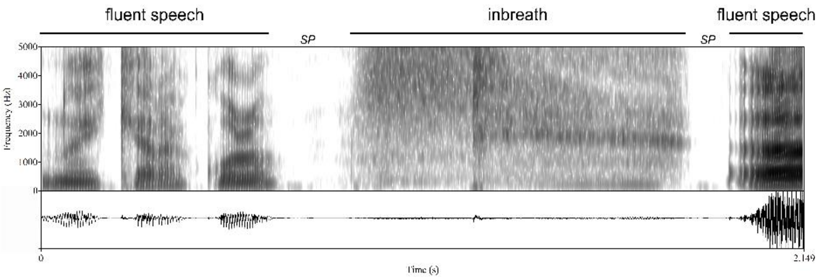


Figure 9. *Inbreath with click (nasal – oral)*
Broadband spectrogram and waveform of the utterance
nu e chiar acolo # (0.181) ¦⇓ (0.988) # (0.099) @ă:
“she is not there yet <silent pause>< audible inbreath with click>
<silent pause> uh”

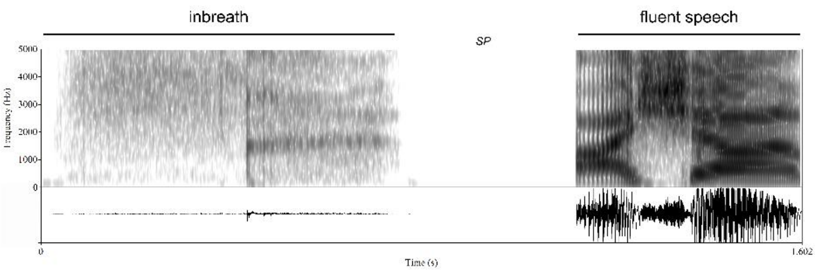


Figure 10. *Inbreath with click (nasal – oral)*
Broadband spectrogram and waveform of the utterance
⇓¦ (0.731) # (0.369) aşa:
“<audible inbreath with click> <silent pause> so”

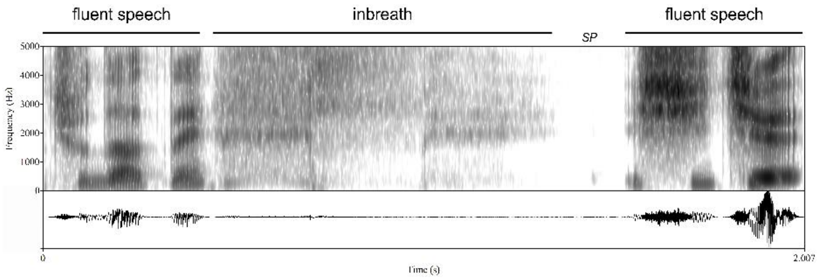


Figure 11. *Multiple inbreaths in one inspiratory breathing episode (oral–nasal–oral). Broadband spectrogram and waveform of the utterance ciudate ⇓⇓⇓ (0.903) # (0.184) şi ce: “strange <audible inbreath> <silent pause> and what”*

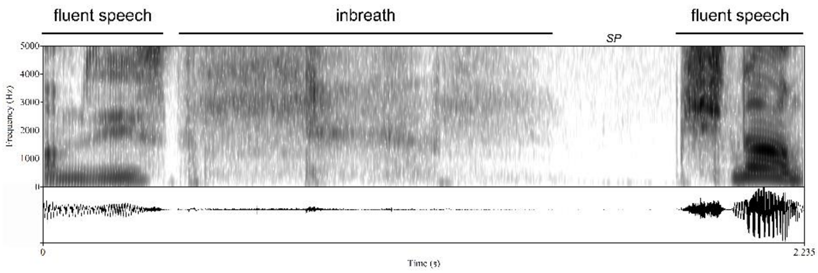


Figure 12. *Multiple inbreaths in one inspiratory breathing episode (nasal–oral–nasal). Broadband spectrogram and waveform of the utterance am zis ⇓⇓⇓ (1.111) # (0.369) şi &na: “I said <audible inbreath> <silent pause> and well”*

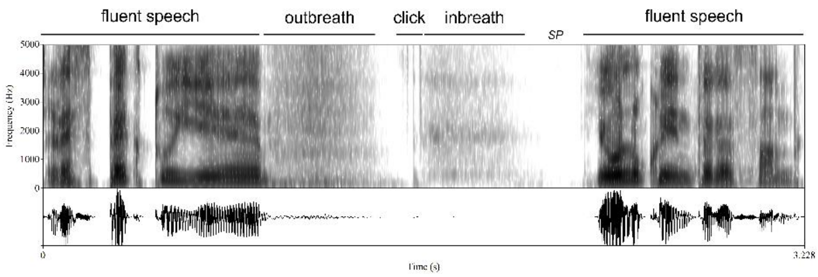


Figure 13. *Outbreath and inbreath*
Broadband spectrogram and waveform of the utterance
respect(l) +e: ⇑ (0.551) ¦ (0.124) ⇓ (0.512) # (0.150) e un lucru interesant
“respect is <non-verbal vocalisations> it is an interesting thing”

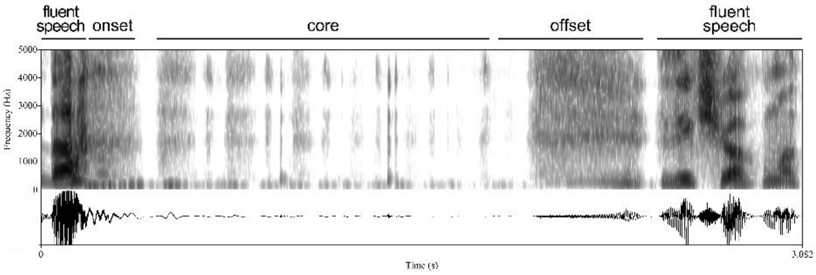


Figure 14. *Inhalation as an offset of laughter*
Broadband spectrogram and waveform of the utterance
da <⇑⇓> (2.309) zic și io acuma
"yes <laugh> I'm just saying"

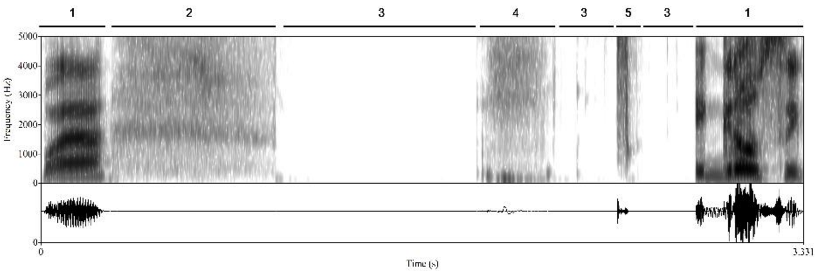


Figure 15. *The interaction between audible inbreaths and other NVVs*
Broadband spectrogram and waveform of NVVs in connected speech showcasing fluent speech (1), deep inbreath (2), silent pauses (3), outbreath (4) and click (5)

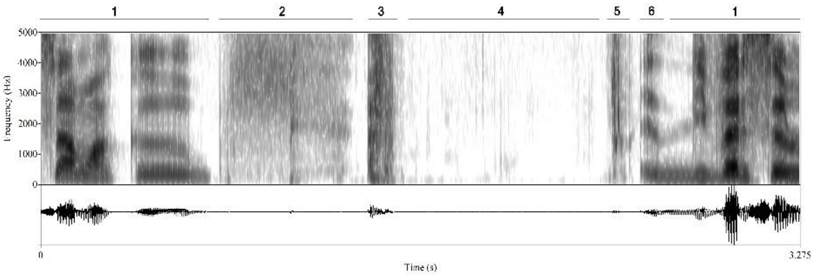


Figure 16. *The interaction between audible inbreaths and other NVVs*
Broadband spectrogram and waveform of NVVs in connected speech showcasing fluent speech (1), inbreath with click (2), cough (3), swallowing (4) followed by a tongue click (5) and a vocalic filler particle (6)

5. CONCLUSIONS

Research on respiratory episodes in Romanian connected speech is still in the early stages. This current article, although exploratory in design, aimed at showcasing the importance of taking into account non-verbal vocalisations in studies related to conversational data. Focusing on breath intakes in human speech, this paper described the differences in tidal breathing as observed in healthy and unhealthy subjects based on previous research on the topic. These breathing patterns were then compared to respiratory kinematics employed in spontaneous speech. An overview of the results obtained from various research fields pertaining to the role of inbreath in speech was presented. The study then proceeded by providing acoustic illustrations of inbreaths in Romanian monologue speech, depicting the various pausal contexts in which breath intakes occur.

Future research on inbreaths in Romanian conversational data can take various directions, ranging from including non-verbal vocalisations as distinct categories in spoken corpora transcription systems and working on time aligned speech based on TextGrids in Praat, to exploring the various functions performed by breath intakes in different communicative settings. In turn, by examining data derived from lesser-resourced languages, we can gain a deeper understanding on the interplay between verbal and non-verbal vocalisations in a cross-linguistic context.

ROMANIAN SPEECH CORPORA

Constantinescu, M.V., G. Stoica, 2020, *Româna ca limbă străină. Corpus* [Romanian as a foreign language. Corpus], București, Editura Universităţii din Bucureşti.

CLRV = Hoarţă Cărăuşu (coord.), 2023, *Limba română vorbită - Corpus şi studii lingvistice* [Spoken Romanian - Corpus and linguistic studies], Iași, Editura Universității Alexandru Ioan Cuza.

CLRVA = Hoarţă Cărăuşu, L. (coord.), 2005, *Corpus de limbă română vorbită actuală* [Corpus of Present-Day Spoken Romanian], Iași, Editura Tehnică, Ştiinţifică şi Didactică Cermi.

CLRVA II (2013) = Hoarţă Cărăuşu, L. (coord.), 2013, *Corpus de limbă română vorbită actuală nedialectală* [Non-dialectal Corpus of Present-Day Spoken Romanian], Iași, Editura Universităţii „Alexandru Ioan Cuza".

CORV = Dascălu Jinga, L., 2002, *Corpus de română vorbită. Eșantioane* [Corpus of Spoken Romanian. Samples], București, Oscar Print.

CORTES = Săftoiu, R. (coord.), 2009, *C O R T E S. Corpus de texte pentru seminarul de limba română contemporană (pragmatica)* [Corpus of texts for the contemporary Romanian language seminar (pragmatics)], Ploiești, Editura Universității Petrol-Gaze din Ploieşti.

Gheorghe, M., S. Măda, R. Săftoiu, 2009, *Comunicarea la locul de muncă. Corpus de interacţiune verbală în mediul profesional* (+ CD) [Communication at work. Corpus of verbal interaction in the professional environment], Brașov, Editura Universităţii Transilvania din Brașov.

IV II (2007) = Ionescu-Ruxăndoiu, L. (coord.), 2007, *Interacţiunea verbală (IV II). Aspecte teoretice şi applicative. Corpus* [Verbal Interaction. Theoretical and applied outlook. Corpus], Bucureşti, Editura Universităţii din Bucureşti.

IVLRA = Ionescu-Ruxăndoiu, L. (coord.), 2002, *Interacţiunea verbală în limba română actuală. Corpus (selectiv). Schiță de tipologie* [Verbal Interaction in Present Day Romanian. (Selective) Corpus. Outline of a Typology], București, Editura Universității din București.

ROVA = Dascălu Jinga, L. (coord.), 2011, *Româna vorbită actuală (ROVA): corpus și studii* [Present-Day Spoken Romanian: Corpus and Studies], Bucureşti, Editura Academiei.

REFERENCES

Atkinson, J. M., J. Heritage (eds), 1984, *Structures of social action: Studies in conversation analysis*, Cambridge, Cambridge University Press.

Bachorowski, J.-A., M. J. Smoski, M. J. Owren, 2001, "The acoustic features of human laughter", *Journal of the Acoustical Society of America*, 111, 1582–1597.

Belz, M., 2023, Defining filler particles: A phonetic account of the terminology, form, and grammatical classification of "filled pauses", *Languages* 8.1: 57.

Belz, M., J. Trouvain, 2019, "Are 'silent' pauses always silent?", *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2744–2748.

Braunschweiler, N., L. Chen, 2013, "Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS", *Proceeding of the 8th ISCA Speech Synthesis Workshop*, Barcelona, Spain, 1–6.

Byon, A. S., 2006, "The role of linguistic indirectness and honorifics in achieving linguistic politeness in Korean requests", *Journal of Politeness Research*, 2, 247–276.

Campione, E., J. Véronis, 2002, "A large-scale multilingual study of silent pause duration", *Proceedings of Speech Prosody*, Aix-en-Provence, France, 199–202.

Charlton, P. H., D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, D. A. Clifton, 2017, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review", *IEEE reviews in biomedical engineering*, 11, 2–20.

Chon, K. H., S. Dash, K. Ju, 2009, "Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation", *IEEE Transactions on Biomedical Engineering*, 56(8), 2054-2063.

Clifton, D., J. G. Douglas, P. S. Addison, J. N. Watson, 2007, "Measurement of respiratory rate from the photoplethysmogram in chest clinic patients", *Journal of clinical monitoring and computing*, 21, 55–61.

Fleming, S., M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko, D. Mant, 2011, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies", *Lancet*, 377, 1011–1018.

Fuchs, S., C. Petrone, J. Krivokapić, P. Hoole, 2013, "Acoustic and respiratory evidence for utterance planning in German", *Journal of Phonetics*, 41(1), 29–47.

Götz, S., 2013, *Fluency in native and nonnative English speech*, vol. 53, John Benjamins.

Jefferson, G., 1978, "Explanation of transcript notation", in J. Schenkein (ed.), *Studies in the organization of conversational interaction*, New York/San Francisco/London, 11–16.

Jefferson, G., 2004, "Glossary of transcript symbols with an introduction", in G. H. Lerner (ed.), *Conversation analysis. Studies from the first generation*, John Benjamins, Amsterdam/Philadelphia, 13–31.

Kienast, M., F. Glitza, 2003, "Respiratory sounds as an idiosyncratic feature in speaker recognition", *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 2003), Barcelona, Spain, 1607–1610.

Kosmala, L., 2020, "On the distribution of clicks and inbreaths in class presentations and spontaneous conversations: blending vocal and kinetic activities", *Laughter and Other Non-Verbal Vocalisations Workshop 2020*, Bielefeld, Germany, 76–79.

Lester, R. A., J. D. Hoit, 2014, "Nasal and oral inspiration during natural speech breathing", *Journal of Speech, Language, and Hearing Research*, 57, 3, 734–742.

McFarland, D. H., 2001, "Respiratory markers of conversational interaction", *Journal of Speech, Language, and Hearing Research*, 44, 128–143.

Mettouchi, A., 2018, "Audible breath intakes in monologues", *Journal of Speech Sciences*, 7(2), 93–106.

Niculescu, O., 2021, "Developing linguistic resources for Romanian written and spoken language", *Proceedings of the 16th International Conference Linguistic Resources and Tools for Natural Language Processing*, Iași, Romania, 21–36.

Niculescu, O., 2022, "Recent tools for teaching and conducting research at the interface between phonetics and phonology in contemporary standard Romanian", in I. Nedelcu, I. Nicula Paraschiv, A. Vasilescu (eds), *Orientări actuale în lingvistica teoretică și aplicată. Actele celui de al 21-lea Colocviu internațional al Departamentului de lingvistică*, București, Editura Universității din București, 145–155.

Niculescu, O., M. Candea, 2023, "Exploring (dis)fluency patterns of identical repetitions in Romanian spontaneous speech", *Proceedings of Disfluency in Spontaneous Speech* (DiSS) *Workshop*, Bielefeld, Germany, 66–70.

Ogden, R., 2013, "Clicks and percussives in English conversation", *Journal of the International Phonetic Association*, 43, 3, 299–320.

Ogden, R., 2021, "Swallowing in conversation", *Frontiers in Communication*, 6, 657190, 19 p.

Pinto, D., D. Vigil, 2019, "Searches and clicks in Peninsular Spanish", *Pragmatics*, 29(1), 83–106.

Rochet-Capellan, A., S. Fuchs, 2013, "The interplay of linguistic structure and breathing in German spontaneous speech", *Proceedings of Interspeech*, Lyon, France, 1128–1132.

Sacks, H., E. Schegloff, G. Jefferson, 1974, "A simplest systematics for the organization of turn-taking for conversation", *Language*, 50, 4, 696–735.

Schegloff, E. A., 1996, "Turn organization: One intersection of grammar and interaction", *Studies in Interactional Sociolinguistics*, 13, 52-133.

Schröder, M., 2003, "Experimental study of affect bursts", *Speech Communication*, 40, 1–2, 99–116.

Sohn, H.-M., 1999, *The Korean language*, Cambridge, Cambridge University Press.

Șuteu, V., 1958, "Arhiva fonogramică a limbii române", *Fonetică și dialectologie*, 1, 211–219.

Torreira, F. J., S. Bögels, S. C. Levinson, 2016, "Breathing for answering. The time course of response planning in conversation", in J. Holler, K. H. Kendrick, M. Casillas, S. C. Levinson, *Turn-taking in human communicative interaction*, Lausanne/ Frontiers Media, 135–145.

Trouvain, J., 2003, "Segmenting phonetic units in laughter", *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 2003), Barcelona, Spain, 2793–2796.

Trouvain, J., 2014, "Laughing, breathing, clicking – The prosody of nonverbal vocalisations", *Proceedings of Speech Prosody*, Dublin, Ireland, 598–602.

Trouvain, J., C. Fauth, B. Möbius, 2016, "Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech", *Proceedings of Speech Prosody*, Boston, MA, USA 31–35.

Trouvain, J., B. Möbius, R. Werner, 2019, "On acoustic features of inhalation noises in read and spontaneous speech", *1st International Seminar on the Foundations of Speech*: BREATHING, PAUSING, AND VOICE, 97–99.

Trouvain, J., K. P. Truong, 2012, "Comparing non-verbal vocalisations in conversational speech corpora", *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, Istanbul, Turkey, 36–39.

Truong, K. P., J. Trouvain, M. P. Jansen, 2019, "Towards an annotation scheme form complex laughter in speech corpora", *Proceedings of Interspeech*, Graz, Austria, 529–533.

Winter, B., S. Grawunder, 2012, "The phonetic profile of Korean formal and informal speech registers", *Journal of Phonetics*, 40, 808–815.

Wright, M., 2011, "On clicks in English talk-in interaction", *Journal of the International Phonetic Association*, 41, 2, 207–229.

Yoon, K.-J., 2004, "Not just words: Korean social models and the use of honorifics", *Intercultural Pragmatics*, 1, 189–210.

Yuan, G., N. A. Drost, R. A. McIvor, 2013, "Respiratory rate and breathing pattern", *McMaster University Medical Journal*, 10, 1, 23–25.