Respiratory patterns and turn-taking in spontaneous Estonian

Inhalation amplitude in multiparty conversations

Kätlin Aare
Respiratory patterns and turn-taking in spontaneous Estonian

Inhalation amplitude in multiparty conversations

Kätlin Aare

Abstract

This thesis explores the relationship between inhalation amplitude and turn-taking in spontaneous multiparty conversations held in Estonian. Respiratory activity is recorded with Respiratory Inductance Plethysmography. The main focus is on how inhalation amplitude varies between the inhalations produced directly before turn onset compared to the following inhalations within the same speaking turn. The results indicate a significant difference in amplitude, realised mainly by an increase in inhalation end lung volume values. One of the possible functions of this pattern is to signal an intention of taking the conversational turn. Another could be a phrasing or grouping function connected to lower inhalation amplitudes within turns.

Keywords
Speech breathing, turn-taking, prosodic boundaries, utterance initiation, respiratory physiology, Estonian spontaneous speech
Respiratory patterns and turn-taking in spontaneous Estonian

Inhalation amplitude in multiparty conversations

Kätiln Aare

Sammanfattning

Denna uppsats undersöker förhållandet mellan inandningsamplitud och turtagning i spontana trepartssamtal på estniska. Andningsrörelser registreras med så kallad Respiratory Inductance Plethysmography. Fokus ligger på hur inandningsamplituden varierar mellan inandningar som föregår samtalsturer och följande inandningar inom samma samtalstur. Resultaten indikerar att samtalsturer föregås av signifikant djupare inandningar än följande yttranden inom turerna, vilket främst är märkbart på lungvolymen vid inandningens slut. En av flera tänkbara funktioner hos detta mönster är som en signal för turbyte, men skillnaden kan även tänkas vara inblandad i frasering och gruppering av turer.

Nyckelord
Andning i tal, turtagning, prosodiska gränser, initiering av yttrande, respirationsfysiologi, estniskt spontaltal
# Table of contents

1 Introduction .................................................................................................................. 1
2 Background ................................................................................................................. 2
   2.1 The respiratory system ......................................................................................... 2
       2.1.1 Lung volume and capacity ............................................................................ 3
   2.2 Speech breathing .................................................................................................... 4
   2.3 Evidence of speech planning ................................................................................ 6
       2.3.1 Turn-organisation and respiratory cues ....................................................... 8
3 Aims and research questions ....................................................................................... 11
4 Method and data .......................................................................................................... 12
   4.1 Speakers ............................................................................................................... 12
   4.2 Experiment procedure .......................................................................................... 12
   4.3 Annotation ............................................................................................................ 14
   4.4 Ethical considerations .......................................................................................... 15
5 Results ......................................................................................................................... 16
   5.1 Data distribution ................................................................................................... 16
   5.2 Lung volume levels in inhalations ....................................................................... 17
       5.2.1 Inhalation start level ..................................................................................... 19
       5.2.2 Inhalation end level ...................................................................................... 20
       5.2.3 Inhalation amplitude .................................................................................... 20
6 Discussion .................................................................................................................... 24
   6.1 Discussion of results ............................................................................................. 24
       6.1.1 Some theoretical implications ...................................................................... 25
   6.2 Discussion of methodology ................................................................................... 26
       6.2.1 Notes on experiment set-up .......................................................................... 26
       6.2.2 Notes on data analysis .................................................................................. 28
   6.3 Future research possibilities .................................................................................. 28
7 Conclusions .................................................................................................................. 31
References ....................................................................................................................... 32
Register ........................................................................................................................... 36
   Figures ......................................................................................................................... 36
   Tables ........................................................................................................................... 36
Appendix: Consent form ................................................................................................. 37
1 Introduction

This thesis is an exploratory study of respiratory activity in spontaneous multiparty conversations held in Estonian and recorded in the Phonetics Laboratory at Stockholm University.

Previous research has shown that speech planning is reflected in respiratory patterns, especially in read speech. For instance, the duration and amplitude of inhalation have been found to correlate positively with the upcoming utterance length in several studies (e.g. Winkworth, Davis, Adams & Ellis, 1995; Fuchs, Petrone, Krivokapić & Hoole, 2013). The location of inhalation is also strongly determined by speech planning. Almost all inhalations in read speech occur at major constituent boundaries, such as paragraphs, sentences or phrases (Conrad, Thalacker & Schönhle, 1983; Grosjean & Collins, 1979).

By contrast, breathing in spontaneous speech shows a less consistent pattern. It has been claimed that as many as 13% of all inhalations in spontaneous monologues occur at grammatically inappropriate locations (Wang, Green, Nip, Kent & Kent, 2010:300), possibly due to the additional demands of real-time speech planning. The effect should be even more pronounced in spontaneous conversation, where the communicative demands are different.

A key characteristic of the conversational rhythm is its oscillating pattern – generally, one speaker at a time has the speaking turn and simultaneous speech tends to be avoided. Thus, the exchange of speaker and listener roles is precisely coordinated by means of turn-taking cues indicating the intention to take, hold or release the turn (McFarland, 2001:128). Breathing patterns have previously been hypothesised to be part of the turn-taking system. Inhalations have been claimed to be an interationally salient cue to speech initiation (Schegloff, 1996:92–93) and to be deeper before turn initiation (Ishii, Otsuka, Kumano & Yamato, 2014:25). Finally, breath holding and exhalation have been suggested as turn keeping and turn-yielding devices, respectively (French & Local, 1983:33–35).

Furthermore, durational properties of respiration have been shown to reflect turn-taking intentions. Speakers tend to minimise pause durations inside the turn by inhaling more quickly, and by reducing the delay between inhalation offset and speech onset (Rochet-Capellan & Fuchs, 2014:12–13; Hammarsten et al., 2015). As inhalation duration and depth have been found to correlate, at least in read speech (e.g. Rochet-Capellan & Fuchs, 2013), the amplitude of non-initial inhalations in a speaking turn should also be smaller.

Previous research indicates that inhaling the approximate amount of air necessary for the upcoming utterance might not be the only purpose of the intake of breath. Inhalations in spontaneous speech might also mark the intention of claiming the conversational floor. Accordingly, the hypothesis under investigation here is that dialogue turns consisting of multiple breath groups should show that the turn-initiating inhalation is larger in amplitude than later inhalations within the turn. Respiratory patterns are collected using Respiratory Inductance Plethysmography (Watson, 1980).

---

1 The author declares no conflict of interest.

2 This research was funded in part by the Swedish Research Council project 2014-1072 Andning i samtal (Breathing in conversation).
2 Background

The background chapter explains the main mechanisms behind the human respiratory system, and introduces the principles that govern speech breathing. In addition, it explains how speech breathing has provided evidence for speech planning and how respiratory patterns have been found to be interactionally motivated.

2.1 The respiratory system

According to Clark, Yallop and Fletcher (2007), most of the respiratory system is contained within the thorax, apart from the upper airways in the subglottal vocal tract. The thoracic cage is made up of 12 ribs which muscle and connective tissue attaches posteriorly to the vertebral columns and anteriorly to the breast-bone (sternum). On the vertical axis, the thoracic cage is limited by the shoulder blades (scapulae) on the posterior side and the collar bones (clavicles) on the anterior side, and by the diaphragm as its base (Clark et al., 2007:168).

Clark et al. (2007) further explain that located inside the thoracic cage are lungs, spungy cone-shaped organs connected to the windpipe (trachea) by two bronchial tubes. Both lungs consist of smaller tubes (bronchioles) ending with tiny air sacs (alveoli). The lungs are connected to the thoracic cage by the pleural linkage, thus forming a single mechanical unit capable of changing the air volume in lungs when the thoracic cage volumes change during the respiratory cycle. Besides performing the vital function of replenishing oxygen and removing carbon dioxide from blood, the lungs provide most of the airflow reservoir necessary for speech production (Clark et al., 2007:168–169).

Clark et al. (2007) describe how the volume of the thoracic cavity is enlarged in two ways during inhalation: firstly, the rib cage is lifted upwards and outwards, and secondly, the floor of the cavity is lowered. The exact balance between the two movements depends on posture, individual habit and respiratory demands. In general, the external intercostal muscles situated between the ribs are responsible for the control of the rib cage dynamics during inhalation. When they contract, the distance between each rib is shortened, raising the rib cage structure, and increasing the thoracic cavity volume. When the diaphragm contracts, it lowers the floor of the thoracic cavity. This action is responsible for inspiratory thoracic cavity changes during quiet breathing. In running speech, the diaphragm has control over increasing the volume of the thoracic cavity for inhalations. Any enlargement of the thoracic cavity results in an increase in lung volume (Clark et al., 2007:169–170).

Reducing thoracic volume, and consequently lung volume, increases the internal air pressure and results in air flowing out of the lungs in order to equalise the internal and external air pressures. The elastic recoil forces set up by the expansion and movement of muscles during inhalation are enough to achieve the necessary lung volume reduction towards relaxation pressure (Clark et al., 2007:170–171). Hixon (1987) proposes that these forces governing both expansion and reduction can be thought of as a spring-like force: if stretched and then released, the spring will rapidly recoil back to its original position (Hixon, 1987:31–33). At lung volumes above the resting level, this relaxation is the result of a passive exhalation generated by positive (i.e. above atmospheric) pressure towards resting level. When lung volumes are below resting level, this process is reversed, as lungs inflate from residual volume to
the resting expiratory level due to the increasing magnitude of subatmospheric pressure (Hixon, 1987:25).

2.1.1 Lung volume and capacity

According to Hixon (2006), volume is one of the key variables controlling respiration. The air displaced by the respiratory apparatus is called lung volume and it corresponds to the change of volume of the thoracic cage. There are four lung volumes, each exclusive of the other, and four lung capacities which are the sum of two or more lung volumes (for more details see Fig. 1) (Hixon, 2006:45).

The tidal volume (TV) is the volume of air inhaled or exhaled during breathing, measured while resting (Hixon, 2006:45). The resting point of the rib cage and lungs is just at the expiratory-end level of the TV and is referred to as the resting expiratory level (REL) (Cleveland, 1998:47). Hixon (2006) reports that the minimum and maximum volumes of air that can be inhaled or exhaled from the tidal levels are the inspiratory reserve volume (IRV) and expiratory reserve volume (ERV), respectively. At the end of a maximum exhalation, the volume of air left in the pulmonary apparatus is called the residual volume (RV) (Hixon, 2006:45–46).

The four lung capacities, as described by Hixon (2006), are as follows. Vital capacity (VC) is the maximum volume of air that can be exhaled after a maximum inhalation. Inspiratory capacity (IC) is the maximum volume of air inhalable from the resting level, functional residual capacity (FRC) denotes the volume of air in the pulmonary apparatus at the resting tidal end-expiratory level, and total lung capacity (TLC) refers to the maximum amount of air in the pulmonary apparatus achievable after a maximum inhalation (Hixon, 2006:46–47).

Although lung capacities vary greatly depending on age, gender, posture and body type, a typical capacity for an adult male is within 5–7 litres of air. In that case, the vital capacity ranges from 3.5 to 5 litres. During quiet breathing, the amplitude of exhalable and inhalable air is around 0.5 litres, and it normally makes up about 10%–15% of vital capacity (Clark et al., 2007:173).
2.2 Speech breathing

Speech breathing commonly refers to the special manner of using the respiratory mechanisms to produce airflow for phonation. According to Euler (1982), speech production usually demands more effort than quiet breathing, and the system is optimised to provide the required airflow. During speech, the rate and volume of inhalation and rate of exhalations are mostly governed by the speech controlling system. For example, this system takes into account requirements for phrasing, loudness and articulations. An important aspect distinguishing automatic or metabolic breathing from voluntary and controlled speech breathing is the brain structure responsible for these mechanisms (see e.g. McKay, Evans, Frackowiak & Corfield, 2003). The first is controlled primarily by the bulboponsine centers in the brainstem, whereas the second also involves cortical structures (Euler, 1982:95–97). The significance of this difference is that the cerebral cortex and other forehead structures control the respiratory system on a higher organisatory level. As other speech functions are also controlled by the cerebral and cerebellar regions of the brain (see e.g. Blank, Scott, Murphy, Warbuton & Wise, 2002), this connects the organisation of speech breathing to other aspects of speech production. Metabolic breathing, on the other hand, is part of the optimal gas-exchange system for life purposes (Euler, 1982:97).

Both the rib cage and abdomen can be used to displace air during speech. Some speakers exhibit the stronger use of rib cage over abdominal contributions, and some speakers show a relatively equal contribution from both the rib cage and abdomen (Hixon, 1982:82). In general, the type of articulation involved, overall vocal effort and the habits of the individual speaker have an effect on the aerodynamic demands of speech on the respiratory system (Clark et al., 2007:172–173).
According to Hixon (1987), speech breathing demands the necessary amount of alveolar pressure to ensure the steady production of utterances. Alveolar pressure is constant during both sustained utterances and conversational speech, but depends on several variables. For example, muscular pressure and relaxation pressure need to be balanced for alveolar pressure to stay constant. More specifically, at high lung volumes, a net inspiratory force is added to the relaxation pressure, but the magnitude of this force decreases as the amount of air in the lungs decreases during speech. At around half of the VC, the net force value is zero. Accordingly, when the level of air in the lungs falls below that, a positive muscular pressure needs to be applied increasingly while the lung volume steadily decreases, but pressure needs to be maintained (Hixon, 1987:27–35; 46–48).

Although speech breathing demands more effort than normal quiet breathing, Clark et al. (2007) observe that both operate in the lower midrange of vital capacity and the minimum respiratory volumes at the end of the exhalation phase tend to be around 30%–40% of VC. However, the tidal peak after inhaling can range from 45% of VC in quiet breathing to 80% of VC in loud speech (Clark et al., 2007:173). According to Hixon (1982), conversational speech is normally encompassed around approximately 40%–60% of VC, while most utterances begin from around twice the resting tidal volume and end just above FRC (Hixon, 1982:82). Hixon, Goldman and Mead (1973) have also compared read speech and conversations and found that regardless of the condition, in most utterances, speech was initiated at 50%–60% VC and terminated at approximately 30%–50% of VC in the upright position (Hixon et al., 1973:93), while a typical speech breathing exhalation phase had the amplitude of approximately 10%–20% of VC, in some cases reaching 30% of VC (Hixon et al., 1973:95). Occasionally, speech ends even lower, in the expiratory reserve level, because speakers aim to finish utterances without inspiratory interruption (Hixon, 1987:44).

The differences between the quiet and speech breathing cycles have been described by Hixon (1987), as follows. While the quiet breathing cycle has been claimed to repeat 12 or more cycles per minute with exhalations lasting slightly longer than inhalations, the frequency of inhalations and exhalations in speech breathing is lower. The relative durations of the phases change because inhalations become considerably shorter than exhalations to minimise interruptions to the speech flow. Exhalations, on the other hand, become much longer due to higher resistances in the upper airway that prevent air from quickly flowing out. Therefore, the patterns of quiet breathing and speech breathing are very different: quiet breathing encompasses relatively equal phases of inhalation and exhalation in terms of duration, amplitude and velocity, whereas speech breathing is characterised by short inhalations and long exhalations (see Fig. 2 for example). As Hixon points out, the hallmark of the volume changes of conversational speech is in fact the irregularity of the breathing cycle (Hixon, 1987:45–46).
2.3 Evidence of speech planning

It is generally agreed that language production is incremental (e.g. Kempen & Hoenkamp, 1987:203–204) – processing one level of information triggers activity on the next level of the production system. In a simplified manner, it is the piece-by-piece process that guides an idea or thought all the way to articulation and thus results in language and speech production, corresponding to specific communicative demands (Ferreira & Swets, 2002:57). However, it is not completely clear how incremental language production exactly is. Some (e.g. Wheeldon & Lahiri, 1997; Levelt, 1989) suggest it is radically incremental in that speakers start articulating when they know the first word of their utterance and during that, the planning of the next phonological word takes place. Others have found evidence that language planning can be more flexible and speakers tend to look for balance between planning and initiating speech quickly, indicating that we are capable of planning larger portions of the utterance beyond the immediate phonological word (Ferreira & Swets, 2002:76–80). That can be exemplified by the use of common expressions, collocations or idioms that tend to form single fixed units in the speakers’ processing memory as opposed to being simply strings of words (see e.g. Wray & Perkins, 2000).

Much of the system for speech planning and organisation has been studied by comparing the proportions and timing of speech and pausing. Pauses in speech have been claimed to be controlled by different variables, such as the rate of speaking, syntactic strength of boundaries, emphatic stress, sentence length, etc. According to Cruttenden (1986), pauses occur at either major constituent boundaries, before words of high lexical content, or after the first word in an intonation group, while the last two types are interpreted as hesitation pauses (Cruttenden, 1986:30–31). Butterworth (1975) studied how strongly pause locations correspond to phonemic clause boundaries and suggests that speakers plan ahead in terms of clauses and sentences but also have the ability to plan superordinate
units consisting of multiple clauses and sentences that form a kind of semantic unit (Butterworth, 1975:84). Krivokapić (2010) reports, based on a study on prosodic phrase length effects on pause durations in read English, that speech is planned quite far ahead beyond the extent of the first phrase but the exact extent depends on the particular speaker (Krivokapić, 2010:3). In addition, Butterworth has pointed out that some pauses could have the communicative function of helping the listener segment the speech (Butterworth, 1975:84). Therefore, pauses not only have the purpose of providing time for cognitive processing to formulate speech, but help guide the listener’s interpretation as well.

Respiratory patterns occurring during read speech have been investigated extensively and a number of respiratory variables connected to speech planning have been looked at closely. Analysis on reading texts has concluded that almost 100% of inhalations occur at syntactic boundaries marked with punctuation or conjunctions (Rochet-Capellan & Fuchs, 2013:1128). More precisely, it has been determined that in read speech, speakers always inhale between paragraphs, very likely inhale between sentences and sometimes also in complex sentences before a comma or connectors (Conrad et al., 1983:224). Remarkably, even reading tasks where speakers are asked to produce only silent, inner speech result in a speech-like respiratory pattern, indicating that breathing is controlled by the cortical structures even when speech is not actually articulated (Conrad & Schönele, 1979:266). Fuchs et al. (2013) have experimented with read German, demonstrating a series of respiratory patterns connected to the syntactic content of the text. For example, they conclude that longer and deeper pauses lead to longer sentences but, at the same time, syntactic complexity does not indicate the same necessity and instead causes more frequent inhalations as compensation (Fuchs et al., 2013:38). Whalen and Kinsella-Shaw (1997) observe a similar effect of utterance length (in terms of durational and syllabic length) on inhalation duration, regardless of whether inhalations were measured acoustically or physiologically (Whalen & Kinsella-Shaw, 1997:145–146). Grosjean and Collins (1979) also investigated the syntactic nature of breathing pauses and set speech rate as a variable. They discovered that at slow and normal rates, speakers prefer to inhale at major constituent breaks, but when the rate is increased, there are fewer breathing pauses and they occur whenever speakers have the need for air. This is caused by the speakers’ wish to minimise the amount of pauses and maximise the speech rate by inhaling very quickly and only when absolutely necessary (Grosjean & Collins, 1979:104–107). Hixon et al. (1973) report similar behaviour from an experiment where speakers were asked to read long sentences at lower lung volumes. They conclude that in those cases, the mechanical aspects of breathing become more important than speech phrasing. According to this study, low levels of air forced the speakers to inhale at unconventional locations in the reading passage to attend to the system’s demands in order to continue the utterance (Hixon et al., 1973:107).

By contrast, spontaneous speech is less predictable than reading – speakers do not have a prepared text to check before producing utterances. Differences in respiratory patterns between read and spontaneous speech have been explored by, for example, Winkworth et al. (1995), who investigated whether the associations known to exist between linguistic factors and lung volumes in read speech also hold in spontaneous speech. They conclude that the location of inhalations mostly (72%) follows clause structure and that longer breath groups have a higher initiation lung volume than short breath groups (Winkworth et al., 1995:132–139). The term "breath group" (Lieberman, 1967:2) itself was introduced to denote the boundaries of a prosodic pattern of simple declarative sentences in normal speech, mainly defined by the fact that they are uttered on a single exhalation. Findings on inhalation locations indicate that speech is mostly structured into breath groups taking into account not only respiratory demands, but also grammatical structure, and, as such, breath groups tend to consist of relatively complete clauses, phrases or sentences. Indeed, grammatically inappropriate inhalations
have been found to occur around 2% for reading and 13% for spontaneous speech in English (Wang et al., 2010:300). However, Winkworth et al. (1995) claim that there is large individual variation to these numbers and that in spontaneous speech, breath groups rather reflect units of meaning (Winkworth et al., 1995:139–140). Rochet-Capellan and Fuchs (2013) investigated spontaneous German in terms of how inhalation depth and duration were connected to syntactic contents of breath groups. Their results show that both the amplitude and duration of inhalation depend on the length of the following breath group, but also whether it started with a matrix clause or some other clause type. If the breath group started with a matrix clause, the preceding inhalation was deeper. In addition, inhalations were found to be deeper when the breath groups contained at least one hesitation (Rochet-Capellan & Fuchs, 2013:1130–1131).

2.3.1 Turn-organisation and respiratory cues

Spontaneous speech usually occurs in the form of a conversation between two or more interlocutors. Conversation, as defined by Jaffe and Feldstein (1970), is a sequence of sounds and silences generated by two or more interacting speakers (Jaffe & Feldstein, 1970:19). A key feature of conversational rhythm is its oscillating pattern – one speaker at a time has the speaking turn and simultaneous speech is generally avoided (Jaffe & Feldstein, 1970:3). Therefore, conversational exchange requires precisely coordinated collaboration in the form of turn-taking movements between the partners – one speaker, who holds the floor, while the other(s) are listener(s) (McFarland, 2001:128). Most studies on spontaneous speech involve two speakers, who exchange turns. Although an undeniably useful source of data for studying mechanisms of turn exchange, the mechanisms used to organise turn-taking become more complicated in multiparty conversations. There are a number of intricate strategies for claiming the conversational floor, even though speakers might not be aware of using these devices. Some strategies are used in order to claim the floor, others to keep the floor by avoiding interruptions, and a third type of devices to hand over the speaking turn to another participant. The signals guiding conversational interpretation are said to be empirically detectable as interactional intentions need to be clearly identifiable during conversations (Gumperz, 1982:159). Listeners are known to turn their attention to stimuli which seem relevant for processing, and, as such, these must be communicated as relevant by the speaker (Wharton, 2009:40). Turn-taking events such as a speaker switch, where one person loses the possession of the floor and another gains it (Jaffe & Feldstein, 1970:19), can be achieved by using certain turn organisatory cues or a combination of them.

The turns speakers take are usually defined as turn constructional units – various unit-types like sentential, clausal or lexical constructions speakers use to construct a turn (Sacks, Schegloff and Jefferson, 1974:702). Feldstein (1973), however, has used the term utterance when determining turn-taking events. According to him, an utterance is made up of sequences of pauses and vocalisations of one speaker that are bounded by switching pauses, where a speaker switch occurs, or vocalisations by other speakers at both ends (Feldstein, 1973:95). The borders of turn-constructional units or utterances can be intensified with the help of a number of prosodic markers, for example, intonation, stress, intensity, voice quality, and the rhythm of phrasing, or pausing and speaking. The extent and manner of use for each feature depends on what the speaker wishes to convey and is affected by the incremental nature of speech: the prosodic content of an utterance is created continuously by moment-to-moment decisions about if and how to continue (Couper-Kuhlen & Selting, 1996:29–30). For example, if the final accented syllable of an utterance in German is said on a mid-level pitch, it is perceived as incomplete, whereas a lowering pitch at the same location would demonstrate the ending of an utterance (Selting, 1995:206). This is connected to declination – a phenomenon whereby pitch lowers during an intonation-group due to a decline in transglottal pressure caused by using up the air
in the lungs (Cruttenden, 1986:162). As such, falling intonation usually signals the end of a sentence or utterance in many of the world’s languages. Another connection here can be made to voice quality: modal voice at the end of an utterance tends to signal incompleteness while irregular phonation or creakiness has been reported to be a phrase-end or turn-end marker due to very low fundamental frequency accompanying it (Slifka, 2007:232). For example, creaky voice as a turn-ending marker is used in Finnish (Ogden, 2001:139–140), Swedish (Carlson, Hirschberg & Swerts, 2005:330), English Received Pronunciation (Laver, 1994:196–197), American English when combined with yeah (Grivičić & Nilep, 2004:8), but curiously not so clearly in Estonian (Aare, Lippus & Šimko, 2014:34). When a speaker’s voice becomes creaky, it can be therefore interpreted as a signal that they have exhausted the air in their lungs and need to inhale soon – providing a convenient location for taking over the conversational floor.

Cues important for the organisation of conversational dynamics can also be inferred from visual signals. The devices people use include gestures, eye-gaze and facial expressions, all of which contribute to the interpretation of speaker and/or listener intentions in conversation. The loss of visual-gestural cues, as happens in phone conversations, has been reported to alter the temporal patterns of interaction: pause durations and stretches of simultaneous speech become shorter (Jaffe & Feldstein, 1970:42). The purpose of spontaneous movements that accompany speech is claimed to constrain the inferential process by triggering a variety of emotion or attitudinal concepts, and altering the salience of linguistically possible alternatives (Wharton, 2009:130; 140–141). According to Bavelas et al. (1995), such interactive gestures can coordinate speaking turns. In fact, they propose that speakers can gesturally take the turn, give away the turn or indicate the floor is free for taking. They also suggest that the words and gestures in spontaneous dialogues are not separate channels but function as a whole (Bavelas, Chovil, Coates & Roe, 1995:397–398).

Respiratory activity during speech can be both visible and audible. Schegloff (1996) has suggested that an audible inhalation functions as a pre-beginning element in turn-taking and projects the onset of talk (Schegloff, 1996:92–93). It is also known that breath holding can function as a marker of turn incompleteness and exhaling can be a turn-yielding device (French & Local, 1983:33–35; Edlund, Heldner & Włodarczak, 2014:36). Before initiating speech, inhalations can be produced with a strong frication or by inhaling in a way that extensively stretches the rib cage to show the intention of speaking with body language. For example, pre-speech inhalatory noise has been found to be audible before short sentences, but single words are usually preceded by silence (Scobie, Schaefller & Mennen, 2011:1784). In addition, research has shown that breathing adapts to dialogue turns and there is some evidence for inter-personal coordination of breathing in turn-taking at a global level (Rochet-Capellan & Fuchs, 2014:5). Rochet-Capellan and Fuchs (2014) have also looked more closely at how breathing cycles might adapt to dialogue events. Their analysis shows that in order to hold a turn, speakers reduced inhalation durations compared to those coinciding with speaker change and thus preserved their turn. They also explain that breathing profiles are different depending on whether speakers are trying to claim the turn or if they are holding the turn: in general, respiratory cycles in turn-taking were longer than in turn-holding and therefore, the breathing pattern of turn-holding was more asymmetrical than for turn-taking. Furthermore, their data on spontaneous German speech demonstrated that turn-taking was more successful after a new inhalation, indicating that speakers coordinate their breathing with turn-taking. (Rochet-Capellan & Fuchs, 2014:12–13). McFarland (2001) investigated the possible influence of turn-taking on respiratory kinematics by comparing the mean inhalatory and exhalatory durations for three breathing cycles directly before and after the onset of speech in scripted dialogues. His results did not reveal significant influence of upcoming speech to
the inhalation duration. However, after the onset of speech, the first inhalation was significantly longer than the following two, which in turn were comparable in duration (McFarland, 2001:136).

By contrast, some contributions to spontaneous conversations do not need to be planned. An example of this is the occurrence of backchannels (Yngve, 1970:568), which are short unplanned listener responses indicating that the listener is understanding and following the speaker (e.g. Heldner, Hjalmarsson & Edlund, 2013:137). Due to the relative unpredictability of backchannels and laughter, they are generally regarded as non-interruptive and not considered as attempts to claim the conversational floor (Heldner et al., 2013:137). As research has shown, vocalised backchannels tend to occur around speaker’s exhalation offset in the listener’s respiratory cycle, and are often located near the onset of listener’s inhalation phase. Occasionally, backchannels also occur in the inhalation phase of the listener and could, in theory, be located almost everywhere in the listener’s breathing cycle (Aare, Włodarczak & Heldner, 2014:50–51). Recent research has also provided evidence that backchannels and other very short utterances, such as short answers to questions (Torreira, Bögels & Levinson, 2015:5) can be produced on residual air. After taking into account the respiratory needs for the upcoming utterance, speakers can choose not to inhale if they already have enough air in their lungs to be able to produce the entire utterance (Włodarczak & Heldner, 2015).

All of these markers are combined with the syntactic content of utterances and help determine the turn-taking intentions of participants. From a practical point of view, the results from studies on speech respiration provide information for applications like human information processing in human-computer interactions. For example, Ishii et al. (2014) have investigated how to predict the next speaker in a multiparty conversation based on the participants’ respiratory patterns. They observed that the person who wants to hold the floor inhales more quickly and with a larger amplitude than the subsequent listeners, and that the new speaker takes a bigger breath than listeners in a turn-changing event (Ishii et al., 2014:23). Similarly, it is known that in a question-answer situation, short replies are mostly produced on residual breath whereas longer responses are preceded by an inhalation (Torreira et al., 2015:7) which can be audible and could help narrow down the possible sequential alternatives in conversations.
3 Aims and research questions

The aim of this thesis is to explore whether speakers use a specific respiratory pattern when claiming the conversational floor in spontaneous conversations. Motivated by the relevant background, the overall hypothesis is that inhalations during speech reflect intentionality. The focus of this thesis is on inhalation amplitudes in uninterrupted speaking turns lasting over multiple inhalation phases.

The main questions of interest are as follows:

i. Does inhalation amplitude of the first inhalation phase of a speaking turn containing several breath groups differ from the amplitude of the next inhalations in the turn?

ii. If inhalation amplitude changes, will it be reflected by inhalation start lung volume level?

iii. If inhalation amplitude changes, will it be reflected by inhalation end lung volume level?

The hypothesis behind question i. assumes that the absolute inhalation amplitude is higher in the first inhalation phase to mark the intention of taking the turn. An alternative explanation could be that there is no difference due to the complicated conversational situation between three participants – each pause in ongoing speech leads to ”re-claiming” the floor if the intention is to continue.

Questions ii. and iii. are an attempt to determine whether changes in inhalation amplitudes are realised by adjusting the inhalation start lung volume level or rather the end lung volume level. It is expected that inhalations will not start at very low lung volumes below the resting level because exhalations reaching so low should generally be avoided due to the physical discomfort. By contrast, inhalation end levels should show a tendency for decreasing in lung volume values if the hypothesis behind question i. is true.

The data at hand (see chapter 4) makes it possible to look into these matters and possibly provide some relevant information for developing the understanding behind speech organisation, conversation structures and reflecting intentionality in speech. The type of data – spontaneous speech – limits the results somewhat, as there are no specifically designed experiments to test the subjects in other conditions as well, but this is partly overcome by comparison with already existing studies. On the other hand, spontaneous speech in the form of multiparty conversations provides a rare type of rich naturalistic data that could reveal much about the behaviour of speakers in natural conversation conditions.
4 Method and data

To carry out research on this topic, four spontaneous three-party conversations were recorded. The language for the conversations was decided to be Estonian for the author of this thesis to be able to distinguish between backchannels and non-backchannels more easily, and participants were chosen accordingly. Detailed descriptions of recording sessions, technical set-up, other principles for the choice of speakers, and annotation rules are provided below.

4.1 Speakers

Twelve participants (8 female and 4 male) took part in the recording sessions. They were all native speakers of Estonian and knew each other relatively well (from a few weeks to 14 years). Apart from two couples living together, the participants described their relationships as being friends. One group of people was living in Stockholm, but used Estonian as their home language and showed no signs of disfluencies that could have been caused by lexical retrieval problems connected to living in another language environment. All others were residents of Estonia and travelled to Stockholm for the recording sessions. The speakers were invited by personal communication and participated voluntarily.

Experiment subjects were chosen primarily by their age and Body Mass Index (BMI). The first is important because aging reduces lung elasticity, elastic recoil forces and the strength of the inspiratory and expiratory muscles (Huber, 2008:323) – therefore ideal participants are relatively young. Secondly, it has been observed that the expiratory reserve volume and functional residual capacity decrease rapidly with an increase in BMI due to the impact on chest wall mechanics (Jones & Nzekwu, 2006:830). Body Mass Index is calculated by dividing mass (kg) with the square of height (m) whereas the normal range is considered to be 18.5–25 kg/m². The speakers’ age ranged from 22–35 (with a mean of 26, SD = 4) years at the time of the recording, and their average BMI was 21.9 (SD = 2.3). They did not report any history of respiratory, speech, hearing or language disorders. The participants had always been non-smokers and were otherwise healthy (no colds, etc) at the time of the recordings.

4.2 Experiment procedure

The recordings took place in the Phonetics Laboratory at Stockholm University. Participants had no knowledge of the aim of the experiment before the recording, and they were completely free to choose their own topic(s) for conversation throughout the recording session. They were given no tasks or guidance on the matter. The speakers were required to fill in a questionnaire (in Estonian, see appendix for the same questionnaire in English) prior to the recording sessions.
Fig. 3. Example of recording for one speaker in LabChart. The first channel depicts sound, the second records expansion of the rib cage, the third level measures the same for the abdomen. The fourth signal is the sum of the changes in lung volume in the rib cage and the abdomen.

Data was acquired on multiple levels at once – respiratory activity was recorded synchronously with audio (see Figure 3) and video. The average length of the conversations was aimed to be approximately 20 minutes.

The participants were instructed to wear tight-fitting clothes to maximize accuracy of the respiratory signals. Respiratory activity was measured with Respiratory Inductance Plethysmography (Watson, 1980). It quantifies changes in the rib cage and abdominal cross sectional area with the use of two elastic transducer belts (Ambu RIP-mate) placed on two levels: one at the level of the armpits and the other on the level of the navel. Both belts were connected to specially developed respiratory belt processors (RespTrack), designed and built in the Phonetics Laboratory at Stockholm University (see Edlund, Heldner & Włodarczak, 2014). They were optimised for low noise and low inference recordings of respiratory movements in speech and singing. A ”zero” button can be used to correct DC offset simultaneously for both rib cage and abdomen signals, and there is no high-pass filter that otherwise causes a decay during periods of breath-holding. The signals from the rib cage and abdomen are weighted with a potentiometer which allows for calibration producing a sum signal estimating total lung volume change. The calibration of the balance between the belts for the estimated lung volume change between the two chest walls was achieved by asking the participants to perform the isovolume maneuvre (Konno & Mead, 1967). To that end, the speaker inhales a small, comfortable amount of air, then holds his/her breath and starts moving the air from the abdomen to the rib cage and back.

In addition to the isovolume maneuvre, the participants were asked to perform two more respiratory maneuvres to measure their resting expiratory level and their vital capacity. For the first, the speakers inhaled some air and then exhaled it without applying any force from their abdominal muscles, until they reached the relaxation level. For the second, they were asked to inhale to the total lung capacity and exhale until reaching the residual volume, while standing straight. These measurements were necessary for determining the minimum and maximum values of the lung volumes under the speakers’ conscious control. All three described tasks and the respiratory signal during the conversation were recorded with LabChart software and PowerLab hardware (both ADInstruments, 2014).
The audio signal was recorded with head-worn microphones using a cardioid polar pattern (Sennheiser HSP 4) in LabChart for synchronisation with respiratory signals and in Reaper software (Cockos Incorporated, 2014) for high-quality sound.

The speakers were asked to stand in a circle around a 1-metre high table for the entire calibration process and recording, keeping their hands placed on the table’s surface to minimise body movements. The participants were recorded in video to be able to check possible movements if the respiratory signals show otherwise obscure patterns. Each speaker was assigned a place to stand at, facing a GoPro Hero 3+ camera (GoPro, Inc., 2013) capable of recording every move from the head to the elbows.

During one recording session, the elastic transducer belt placed of the rib cage on one participant moved drastically from its initial position. The data from that speaker has been used in the analysis only up until the respiratory signal was clear enough. Once the pattern became flatter and irregular, the information became excluded. The speaker readjusted the belt to its initial position, but at that point the signal was already compromised and not suitable for further analysis. This happened approximately halfway into the conversation. However, the recording continued in order to retrieve more data from the other two speakers.

4.3 Annotation

Annotation was done using Praat (Boersma & Weenink, 2015). The audio signal was annotated manually and labels were named accordingly to what was happening in the conversation: speech stretches, backchannels, laughter, coughing, pauses. Since backchannels can often correspond to lexical material produced as legitimate answers to questions (e.g. jah, ’yes’), backchannels were determined based on the particular communicative situation. If the listener produced a short response functioning as an answer to a question, it was not counted as a backchannel. If, however, the listener produced a short response simply to give feedback to the listener without being asked anything, it was categorised as a backchannel. The reason for separating backchannels from the rest of speech was to be able to exclude interrupted speaking turns later in the analysis. As the listener’s purpose of producing backchannels is not to interrupt, but rather to reflect on the information given by the speaker, backchannels are counted as essentially non-interruptive elements in the conversations. The same status was given to laughter and coughing. Finally, pauses shorter than 200ms were not marked in the data, as that is considered the minimum duration of a silent pause (Laver, 1994).

The respiratory signal was annotated automatically with a Python script and then corrected manually due to some inaccuracy in placing the positive and negative peaks of respiratory cycles. A total of 11.9% of the interval borders were either moved or added. Even after manual correction, there were speech stretches coinciding with the ends of inhalation phases and each such occurrence was excluded from the analysis.

The data was analysed with another Python script to determine turn-taking events and their positions in the conversations. Following Jaffe and Feldstein (1970), silences and overlaps of speech were classified depending on whether they coincided with speaker change or were followed by more speech from the same speaker. Accordingly, the speaking turn has been defined as an uninterrupted series of

---

7 The Python and Praat scripts mentioned in this section are in a large part credited to M. Włodarczak and are available upon request. The Praat TextGrid processing and analysis toolkit used on this data set is described in greater detail in Buschmeier and Włodarczak (2013).
speech segments from a single speaker. While Jaffe and Feldstein (1970:19) regard any vocalisation as possession of the floor, here the purpose of the vocalisations has also been considered when determining turn-taking events, and backchannels are treated differently from the rest of speech. As stated above, backchannels, laughter and coughing were not categorised as elements produced to claim the conversational floor. Consequently, the division into between-speaker and within-speaker silences and overlaps was carried out by assigning laughter, backchannels and coughing the same status as silence. Following Heldner, Edlund, Laskowski and Pelcé (2009), a between-speaker silence is a pause between speech produced by two different speakers, whereas a between-speaker overlap is simultaneous speech produced by different speakers, resulting in a speaker switch, i.e. a change in who possesses the conversational floor. Accordingly, a within-speaker silence is a pause during a single speaker’s speech, while a within-speaker overlap is a short stretch of speech uttered simultaneously to another speaker’s speech, not resulting in a speaker switch (for more detail, see e.g. Heldner et al., 2009, or Heldner & Edlund, 2010).

Finally, a Praat script was used to find the uninterrupted speaking turns that contained several breath groups. These turns were further limited by laughter produced by the speaker – as laughter typically ends in very low lung volumes, the following inhalation amplitude tends to be very high simply to achieve a balance point again. Therefore, all turns preceded by or including laughter were excluded from the analysis. After that, the same Praat script was used to locate inhalation start and end points of the remaining speaking turns and measure corresponding lung volume values.

The values were then normalised with respect to the minimum and maximum lung volume values measured during the calibration phase. The minimum values of vital capacity proved not to be sufficient for calculating the normalised lung volumes for all of the speakers, as some of them failed to reach the lowest possible lung volume levels. Therefore, for some informants, these measurements were taken from a later point during the conversation to ensure a positive normalised value.

### 4.4 Ethical considerations

The participants gave their written consent prior to the recording sessions. The speakers were informed of their right to withdraw from participation at any given moment during the recording session, and of their right to cancel the authorisation of using collected data even after participation. They were also aware that their personal information will be disclosed and the data collected will only be used for research and teaching purposes.
5 Results

The results presented here are based on statistical analysis carried out with R (R Core Team, 2015) after thorough limitations applied to the raw data (see previous chapter for more details). General information about the retrieved data is provided in the next section. Further results on inhalation amplitude are demonstrated later on in this chapter in the following order: inhalation start level, inhalation end level, and inhalation amplitude.

5.1 Data distribution

The data was collected over four recording sessions from 12 different speakers. Conversations lasted for approximately 22 minutes (19 to 26 min), totalling up to 1h and 28 min of conversation time. After excluding all interrupted speaking turns, stretches of speech coinciding with inhalations, stretches of speech preceded by or including laughter by the same speaker, and all single-breath-group turns, 50 suitable speaking turns with the total duration of 8 min and 15 s were left for analysis. These turns consisted of 128 breath groups and were produced by 11 participants. By contrast, the speakers produced 761 turns (including simultaneous speech and interrupted turns) that consisted of a single breath group, and therefore, the amount of turns analysed here is merely 6% of all recorded speaking turns (see Fig. 4 for comparison).

![Bar chart showing the distribution of multi-breath and single-breath groups.](image)

Fig. 4. Number of turns consisting of multiple and single breath groups.

As can be seen from Figure 5 (below), the speakers’ individual contribution to the final data set analysed here shows great variation, from 0 to 33 breath groups out of the total of 128.
Fig. 5. Number of breath groups per speaker left for analysis after filtering raw data.

Each breath group corresponds to one inhalation phase and each analysed speaking turn consisted of at least two breath groups. However, there were some turns that were significantly longer than only two breath groups. A fifth of the turns included another, third inhalation phase. Remarkably, there was one speaking turn which contained 9 inhalation phases. This information is depicted in Table 1 (below).

Table 1. Inhalations per speaking turn

<table>
<thead>
<tr>
<th>Number of inhalations</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4 ≤</td>
<td>9</td>
</tr>
<tr>
<td>Σ</td>
<td>50</td>
</tr>
</tbody>
</table>

For the purposes of the following analysis in terms of the number of inhalations produced in each turn, focus will be on the first two inhalation phases. Those will be the turn-initial inhalation and the first turn-medial inhalation. This choice is motivated by the very limited amount of conversational turns consisting of more than two breath groups.

5.2 Lung volume levels in inhalations

This section addresses inhalation amplitude in greater detail. Firstly, focus will be on normalised speaker-specific lung volumes (see Table 2 and Figure 6). Secondly, this section looks at the start and end levels of vital capacity used before turn onset and within the turn. Finally, section 5.2.3 will demonstrate the patterns of respiratory amplitude detected in these conversations.
Table 2. Speaker-specific normalised lung volume values (% VC). The table is based on normalised inhalation start and end points, and normalised inhalation amplitudes. Average values are calculated based on all inhalations a speaker took, regardless of whether it was the first or one of the following inhalations within the turn. "NA" = not available.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (inhalations)</td>
<td>8</td>
<td>2</td>
<td>33</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>27</td>
<td>15</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Average start level (% VC)</td>
<td>27.1</td>
<td>27.7</td>
<td>26.3</td>
<td>12.4</td>
<td>27.5</td>
<td>18.4</td>
<td>10.7</td>
<td>NA</td>
<td>35.7</td>
<td>38.0</td>
<td>14.2</td>
<td>40.3</td>
</tr>
<tr>
<td>Average end level (% VC)</td>
<td>37.8</td>
<td>46.9</td>
<td>42.8</td>
<td>30.4</td>
<td>41.6</td>
<td>34.3</td>
<td>32.6</td>
<td>NA</td>
<td>61.4</td>
<td>57.3</td>
<td>24.1</td>
<td>46.5</td>
</tr>
<tr>
<td>Inhalation amplitude (% VC)</td>
<td>10.8</td>
<td>19.2</td>
<td>16.5</td>
<td>18.1</td>
<td>14.1</td>
<td>15.9</td>
<td>21.9</td>
<td>NA</td>
<td>25.6</td>
<td>19.3</td>
<td>9.9</td>
<td>6.1</td>
</tr>
</tbody>
</table>

Inhalation amplitude was measured on the basis of two points: the inhalation start lung volume level and the corresponding inhalation end lung volume level. These values were normalised for each speaker using their VC values. Figure 6 shows the normalised individual inhalation amplitudes used by each speaker (also shown in Table 2) throughout the automatically detected multi-breath-group speaking turns.

Fig. 6. Speaker-specific normalised inhalation amplitudes (% VC). Speaker 8 is not depicted on the figure as she did not produce any (0) adequate breath groups and turns.4

4 Here and on next figures showing boxplots, the vertical line inside the boxplot refers to the second quartile, i.e. the median value. The limits of the boxes reflect 50% of the data, ranging from 25% to 75%, and the ends of the “whiskers” indicate the minimum and maximum values of the samples, excluding outliers.
The information on Figure 6 is based on the varying amount of overall data. For example, speakers 2, 6, 8 and 12 produced 0-3 breath groups, whereas the rest of the speakers contributed with more than 6 breath groups. However, the graph shows that most speakers used a relatively uniform range of their lung capacity for inhalations before speech onsets. The normalised average inhalation amplitude is 19% of vital capacity ($SD = 11.9$).

### 5.2.1 Inhalation start level

The inhalation start level was measured for turn-initial and first turn-medial inhalation and normalised with respect to the VC range used by the speaker. The results are presented in Figure 7.

![Figure 7](image)

**Fig. 7.** Normalised inhalation start levels (% VC) according to no. of inhalation. "1" marks the turn-initial inhalation, "2" marks the first turn-medial inhalation.

Figure 7 shows a relatively equal range and median of the starting inhalation level for both inhalation locations, although a slightly wider range can be observed for the second inhalation. The corresponding median values of inhalations do not show a tendency of lowering towards the end of the turn but stay in the vicinity of 30% of VC. The mean lung volume values are also of comparable sizes: 27.4% for the first and 27% for the second inhalation phase ($t(98) = .171, p = .864$). The average inhalation start level over all speakers and inhalation locations (in this case including inhalations after the second per turn) is 26% of VC ($SD = 11.4$). The range in which the first inhalations were started is 4%–46%, and 1%–52% of VC for the second inhalation.
5.2.2 Inhalation end level

The inhalation end levels have been normalised similar to inhalation start levels. The results for each inhalation location are presented on Figure 8.

![Box plot of normalized inhalation end levels](image)

*Fig. 8. Normalised inhalation end levels (% VC) according to no. of inhalation. “1” marks the turn-initial inhalation, “2” marks the first turn-medial inhalation.*

Figure 8 shows very different values compared to those in Figure 7. As expected, the values for inhalation end levels are higher than inhalation start lung volume values, and range from around 30% to 55% of VC with an average of 45% (SD = 15.4). The graph indicates a lowering effect for the median for the second inhalation in a turn. The mean values for the inhalation end lung volume were 49% and 42%, respectively. The difference in the means of the two inhalations is statistically significant (t(98) = 2.854, p = .005). The first inhalation phase end values ranged from 19% to 81%, and the second inhalation end values ranged from 10% to 77% of lung volume.

5.2.3 Inhalation amplitude

As might be expected given the results in previous sections, inhalation amplitude with respect to the location of the inhalations shows some variation. Inhalations no 1 and 2 were measured for each turn. The results are provided in Figure 9.
Figure 9 shows that the first inhalation is greater in amplitude than the following inhalation phase. More precisely, the boxplots demonstrate that the first inhalation is larger in range and has a higher median. The means of the first and second inhalation amplitudes equal 21% and 15%, respectively. This difference is statistically significant ($t(98) = 2.854, p = .005$). The normalised lung volume range of the first inhalation amplitude varies is 2.3% to 50.7%, whereas the second inhalation amplitude varies in the range of 2% to 72.4% of normalised lung volume.

For an easier comparison, the normalised start and end values have been put side-by-side in Figure 10.
Fig. 10. A comparison of the normalised start and end lung volume levels (% VC) for the first two inhalation phases of a turn. All data is normalised to the speakers’ maximum and minimum VC values. “1” marks the turn-initial inhalation, and “2” marks the first turn-medial inhalation. “start” or “end” are marked accordingly to the point of measurement – either the start or the end of the inhalation phase.

Although the results on this graph have already been introduced, Figure 10 makes it apparent that the difference in amplitude comes from a higher inhalation end value of the first phase. This is represented by all measured values on the graph: the median as well as the highest and lowest lung volume values. The second inhalation’s end value is lower in all of these measures.

A more detailed relationship of the start and end lung volumes of the inhalations can be seen in the scatterplot in Figure 11.
Figure 1 shows a positive relationship (correlation: $r(48) = .657$, $p = .000$) between the inhalation start and end lung volume levels for both inhalation phases. The graph indicates that inhalation phases that start on a low lung volume level, end on a relatively low lung volume as well, and analogously for higher values: the inhalations beginning at high lung volume values also end at higher lung volume values.
6 Discussion

The discussion chapter will analyse the findings in the light of already existing work. In addition, this chapter aims to provide notes on the methodology used in the experiments and working with the data by taking into account some shortcomings and possible improvements. Finally, a number of future developments applicable to this research topic are introduced.

6.1 Discussion of results

While the data analysed for this study reveals significant differences, it is important to emphasise that the conclusions made from these results cannot be too definitive. The amount of data available for analysis was very limited due to the small number of recordings and the thorough filtering process applied before the analysis. Other factors might have limited the amount of usable data further, although less directly. For example, it has been claimed that communicative behaviour is less dependent on language, whereas the correlation between culture and communication is much stronger (see e.g. Wierzbicka, 1998). By comparing Finnish and Estonian radio dialogues, it has been shown that pause durations and turns taken in Estonian conversations are half as long as the Finnish ones. Finnish conversations have been described as polite - listeners quietly paying attention to what the speaker is saying with rarely occurring interruptions. Estonian conversations, however, exhibit frequent interruptions of each other’s turns, giving way to prompt turn changes. The same study also reported that Estonian speech is much faster than Finnish (158 words per minute vs. 114 words per minute, respectively). (Pajupuu, 1995, as cited in Ryabina, 2008:42). This could help explain why most turns analysed here only consisted of two breath groups. Considering the small number of recordings, the measured amount of even two breath groups per turn from typical Estonian (multiparty) conversations seems a relatively satisfactory result.

The normalised inhalation start values for both inhalations measured in this study showed a relatively uniform lung volume range, although the second inhalation phase of a turn exhibited slightly more variation. The lung volume level from which most measured inhalations started was approximately 27% of VC. In general, inhalations tend to start in the lower midrange of VC and have been reported to start form around 30%-40% VC (see section 2.2 for a more detailed description). The obtained values in this case are slightly lower than expected. This could be explained by a number of factors. Firstly, it is likely that speakers did not reach the real maximum inhalation and exhalation points during the calibration stage, affecting the normalisation and resulting in shifted values. Secondly, some speakers might normally rely on rib cage breathing while speaking, but since they were asked to use both the rib cage and abdomen for the calibration, the actual zero point could have been different. Thirdly, the zero point of the two lung volumes might have been assigned incorrectly during the calibration process due to a too forceful or weak sigh produced by the speakers. However, a small range at this level also shows that speakers were not inclined to let exhalations end below a certain point which is probably very close to the relaxation level. The results thus indicate that speakers started inhaling before exiting the physically comfortable lung volume range, although with some exceptions.
The normalised inhalation end lung volume levels demonstrated a notably different pattern. The inhalation preceding turn onset ended at a significantly higher lung volume level than the following inhalation located inside the speaking turn. More precisely, the first inhalation ended on average approximately at 49% of VC ($SD = 14.5$) while the second ended around 42% of VC ($SD = 15.4$). These values are 15%–22% higher than the inhalation start values. Due to the already mentioned reasons, these values might not fall onto the exact suggested lung volume level and could be shifted to a lower value, they are most likely still located in the lower half of the inspiratory reserve level, as would be expected.

Correspondingly, most of the normalised inhalation amplitudes fell in the range of 15%–21% of VC, with the occasional minimum amplitudes at around 2% of VC and the maximum amplitudes covering over 50% of VC. The average speaker-specific inhalation amplitude was 19% of VC ($SD = 11.8$). The turn-initial inhalation amplitude had the average value of 21% of VC ($SD = 10.4$) and the first turn-medial inhalation 15% of VC ($SD = 11.8$). The difference between them is statistically significant and is due to the difference in inhalation end lung volume values. The amplitude between inhalation start and end lung volume values corresponds to the overall expected difference, also reported by Hixon et al. (1973:95, discussed here in section 2.2).

Overall, the results obtained were able to show that the first inhalation phase is indeed larger in amplitude than the following inhalation phases. Although the number of occurrences for the third inhalation phase of a speaking turn was too small for analysis and therefore not presented here, the normalised lung volume ranges of the third inhalation phase did show a tendency for further lowering of the overall range, mean and median of the corresponding lung volume amplitudes, as well as the start and end values of the inhalations. This indicates a potential lowering trend of lung volumes to some extent over a speaking turn, possibly guided by a rough estimate of how much is left to be said during a turn. In addition, the larger turn-initial inhalation amplitudes might suggest that the difference itself could identify the intentions of taking the turn or holding the turn. However, it could also be that increased inhalation end values are responsible for signalling the intention of taking the turn instead of the overall difference in amplitudes. An inhalation ending at higher lung volumes is simply physically more visible than a greater amplitude realised by starting at a lower lung volume level, and could therefore signal speakers’ intentions in a more perceptually salient manner.

### 6.1.1 Some theoretical implications

The results further indicate that using specific respiratory patterns can alter possible interpretations for utterances. This is a common feature of natural behaviour accompanying language use, and also of prosody in general (Wharton, 2009:141). Lehiste (1970) has defined prosody as comprising suprasegmental aspects of speech such as syllable structure and intonation with their acoustic counterparts measurable as fundamental frequency, duration and intensity (Lehiste, 1970:1–3). However, it is also possible to define prosody by its function which is to group speech stretches rhythmically, assign emphasis, convey speech acts, organize turn-taking, reflect the speaker’s emotions or attitudes, and so on (see e.g. Ladd, Scherer & Silverman, 1986; Cutler & Pearson, 1986; French & Local, 1986). Wharton (2009) argues that prosodic inputs are highly context-dependent and can vary on a scale from non-verbal to verbal, or more naturalistic inputs to those of properly linguistic, as both ends of the scale influence the interlocutor’s comprehension process (Wharton, 2009:140–141).

According to that, even non-verbal devices such as respiratory patterns and their communicative influence could be part of several prosodic categories. For example, the division of the speech flow
into breath groups undeniably applies a certain rhythm and tempo to speech. Another example is the use of breathing to signal the wish to take the turn, hold the turn or release the turn, which can markedly change the dynamics of conversations. Breathing can also reveal the speaker’s or listener’s emotions and attitudes (see e.g. Bloch, Lemeignan & Aguilera, 1991) and has been linked to the term "inarticulate" prosody (Sabbagh, 1999). Therefore, respiratory patterns during speech are relevant from a broader interactional point of view. The results of the thesis at hand provide even more evidence in favour of this.

Based on the relevance theory, interpreting the actual linguistic representation of utterances is not enough and human cognition is set to search for additional relevant information to make communication more efficient (Sperber & Wilson, 2002:3–4). Speech breathing can offer such information on all three levels of the contextualization cues proposed by Gumperz (1992): conversational management in turn-taking, sequence organization and disambiguation of speaker intentions, and generating interactional atmosphere (Gumperz, 1992:130–152). The suggested turn-claiming function of the first inhalation phase of a turn with multiple inbreaths might not be the only turn-organizatory cue used to make the speaker’s intentions clearer. There is a possibility that speakers opt for using multiple cues at once to disambiguate their intentions more quickly and with less likelihood of interruptions. That could be achieved by combining short and shallow inhalations, which are potential locations for turn switches, with perhaps stronger frication noise to signal that the speaking turn is not finished (the connection between respiration and turn-taking is explained in section 2.3.1).

According to Zipf (1949), all human action, including linguistic and communicative behaviour is governed by the principle of least effort. In brief, it is motivated by the fact that speakers would prefer to forward their message with as few communicative means as possible, whereas listeners would prefer to use the least amount of effort to infer the meaning behind the speaker’s message correctly. The conflict of this situation is overcome by communicating at a balance point of producing the least effort to achieve minimally sufficient clarity (Zipf, 1949:19–22), or in other words, minimally sufficient discriminability (Lindblom, 1990:404–405) in a context. As such, the use of inhalation amplitude or inhalation end value as a turn organizing cue would imply that it is a device that optimises conversational interaction enough to balance the extra strain put on the respiratory system.

6.2 Discussion of methodology

6.2.1 Notes on experiment set-up

The entire set-up of the recording session is important for achieving useful results. Therefore, each detail needs to be taken into account. For example, instructing the participants to wear tight-fitting clothes helps reduce the noise in the recorded signal. This condition makes sure that the RIP-belts are not placed on clothing that can move freely around the body, and thus minimises the possible effect clothing might have on the results. Another aspect to emphasise is the posture of speakers. On the one hand, the recording session is not comfortable – the speakers are required to stand through the entire calibration session as well as for the actual conversation part, lasting up to an hour depending on the recording session. This can be relieved in a very small part by having the participants keep their hands on the table but the whole recording session can still be tiresome for some participants. The calibration of the belts might end up somewhat shifted due to changes in posture during the session. On the other hand, without these requirements, the quality of the respiratory signal could be considerably more
compromised. For example, during the very first pilot studies in this area of research in the Phonetics Laboratory at Stockholm University, the speakers were asked to sit during the recording sessions to make them feel more comfortable. This resulted in a substantial loss of initial posture as it is very easy to slowly fall lower and lower on the chair, also causing the belts to shift from their assigned position when the participants’ shirts are pulled upwards due to the body movements in the chair. Even if the speakers were asked to pay attention to this and prevent it, the movement is barely noticeable over a longer time stretch. Therefore – although not ideal, it is important that the speakers are recorded while standing and that the calibration phase is completed as quickly as possible to make the whole session less strenuous for the participants.

The situational difference between the calibration process and actual conversation time might have had some influence on the speakers. This possibility was mostly visible from assigning the zero point during calibration. While the speakers exercised much control over their posture and paid a lot of attention on how to breathe with their rib cage and abdomen for the respiratory manoeuvre tasks, this controlling effect was lost after the start of the conversation. Although the zero value of the belts is generally assumed to be assigned correctly, the actual zero point could have moved slightly since the speakers were, for example, used to breathing mainly using their rib cage during speech and applied that style of breathing for the remainder of the session. However, this was possible to fix easily by normalising the values of the respiratory signals to the corresponding minimum and maximum points of vital capacity. Occasionally, the minimum point needed to be measured from another location during the conversation because the speakers failed to reach the minimum values during calibration. This was detected from the normalisation phase of data analysis due to some values staying negative regardless of normalising the values in the range of [0; 100]. The alternative minimum lung capacity values were the ends of laughters, since a laughter is very likely to end extremely low in terms of lung volume. Whether the speakers actually reached the minimum and maximum lung volume values is unknown, but given the unexpectedly low lung volume levels used for speech breathing in this study, the speakers most likely did not reach those points during the calibration. This should not play an important role in the data analysis as the effects are consistent and relatively clear even though the minimum and maximum lung capacity values could be more accurate. Nonetheless, participants should be encouraged to put more muscular effort into achieving to reach these points by using more precise instructions. An alternative for using the maximum inhalation and exhalation values for normalisation is to use tidal breathing instead. That would require measuring the minimum lung volume values during quiet breathing to assign a possibly more accurate zero point at the resting level.

As for the possible effect of the recording situation on making the conversation feel artificial for the speakers – it showed no relevant influence on the conversations. The participants were indeed in unfamiliar surroundings, wearing two transducer belts around their bodies, had microphones attached to them, several cords hanging from the devices, and a camera facing each of them throughout the entire calibration and recording time. However, the speakers seemed to feel relatively comfortable. This was achieved mainly with the calibration session prior to the actual recording of the conversation. The calibration required respiratory tasks and movements which were unusual to perform, but somewhat amusing to witness while their friends were asked to do them. Therefore, the initial awkwardness went over already before starting the conversations and the speakers seemed to act quite naturally, except for the continuous requirement of standing still and not producing too many gestures.

Although being in unfamiliar surroundings under recording conditions might influence speakers to act more politely towards each other, they were also very aware of the fact that besides the person overseeing the recording from behind the curtains, they were alone in the room. This and losing the
awareness of the recording situation during the calibration session most likely gave the speakers a certain comfort-zone where they could act as they normally do when engaging in conversations. Furthermore, most spontaneous conversations analysed in existing studies have had the form of a dyad – with only two speakers. By contrast, in this study, the conversations were expected to be more naturalistic and complicated due to the larger number of participants, as attention is drawn more to the conversational partners than to other aspects connected to the recording itself.

None of the speakers used the opportunity to withdraw from participation in the middle of the recording session. Likewise, the participants did not complain about the conditions of the experiment and have not reported any need to withdraw data recorded from them.

6.2.2 Notes on data analysis

The exclusion rules used to filter raw data into a useful data set in order to find uninterrupted turns consisting of multiple breath groups could be advanced even further. If a larger data set would allow this, pause durations could be used for limiting the data. If turns that consist of multiple breath groups have a very long inhalation pause somewhere in the middle of the turn compared to most inhalation pause durations, it could serve as a marker for topic change. Marking topic changes is complicated and falls more into the area of conversation analysis (see e.g. Schegloff, 2007). However, it could serve as an important addition to clarifications if data shows more occurrences of turns consisting of considerably more than two breath groups, as witnessed in this data set. Long pauses within turns might also imply that the turn is released and then claimed again – perhaps nobody else has taken over the speaking turn and the last speaker therefore continues to make it either more explicit that the turn should go to somebody else, or starts another topic. In the latter case, it is unclear if or how the inhalation amplitude patterns would change if topic would be changed after an extended pause during a single speaker’s turn, and what these patterns might indicate. Furthermore, to focus on intentionality behind speech production and turn-taking more specifically, another limitation could be set on speech onset time after the start of the corresponding exhalation phase. The longer the speech onset time is, the less the inhalation preceding the speech stretch can reflect on intentions behind it.

Given the relatively small amount of data used for this study, it was possible to work very closely with raw data. Manual correction of automatic annotation was a practical necessity to maximise the amount of useful data but in case of a larger data set, this could be avoided. The extent of how much manual annotation is realistic on a bigger data set would depend on the number of additional recordings. For instance, manual correction of assigning the minimum and maximum peaks of the respiratory cycle could be avoided completely, although some useful information would be lost with this compromise. However, in terms of timing in the breathing cycle, it is not distinguishable whether a very short utterance (VSU, shorter than 1 sec) is a backchannel or a short answer to a question (for more detail about VSUs, see Edlund et al, 2010, and for Estonian results on backchannel location in the respiratory cycle, see Aare, Włodarczak & Heldner, 2014). Precise division requires manual classification of each VSU, which is important to determine if a VSU is intended to claim the conversational floor or not and so far there are no reliable means of doing this automatically.

6.3 Future research possibilities

There are a number of directions in which the current research questions discussed in this thesis could be expanded. Some of them are introduced briefly in this section.
Regarding inhalation amplitude and turn taking, a factor that has gained too little attention is backchannel communication. Even if backchannels are considered non-interruptive, it could be that any kind of vocalisation produced by listeners influences the speaker. The results in this thesis show a relatively big variation in normalised inhalation end lung volume values in the second inhalation phase and some of this variation could be explained by influence from speaking partners. Therefore, the inhalation pauses inside a turn following a listener’s backchannel could be investigated further to see if and to what extent these contributions have an influence in increasing the next inhalation amplitude of the speaker.

As it has been revealed by previous research, inhalation amplitude and duration are correlated positively. If each next inhalation in a speaking turn is smaller in amplitude, the durations of inhalations should also show a decreasing tendency. The results presented here only focussed on the first and second inhalation phase of a turn but, as mentioned in the results discussion, the third inhalation phase showed a general tendency of having an even smaller amplitude and a lower median than the previous inhalation. As the number of occurrences was too small, the details were not included in the results. However, more data could demonstrate that this is a stronger pattern. The motivation behind this could be that it is another way to ensure continuing the speaking turn – the longer it lasts, the more likely it could be that other participants would interrupt. Therefore, minimising inhalation durations and inhaling slightly less each time might serve the purpose of holding the conversational floor. If that is the case and the values tend to decrease, there is still a limit as to how often inhalations can occur from the points of both minimum potential interruption locations and the physical comfort of the inhalation frequency – not to mention the requirement of providing sufficient clarity in terms of phrasing and grouping, which are known to guide the listeners’ interpretation process.

A feature of breathing that might reveal even more about turn organization is respiratory noise. Measuring the acoustic intensity of frication during inhalation pauses could reveal a systematic behaviour connected to turn-holding. Speakers might exhibit more frication noise while getting closer to the end of their intended turn. This could be motivated by the same reasons as the increase in inhalation amplitude and duration – it could function as a turn-holding device. An increase in inhalatory noise would make the pause perceptually less prominent and therefore an inconvenient moment for an interruption, since sound is being sustained during the pause. A combination of these cues or all of them together together could be used to intensify the speaker’s intentions and for making the interpretation process perceptually more salient for the listener(s). Furthermore, future studies could be able to show where speakers intend to take, hold or yield a speaking turn but fail to do so. Research on the intentions of the speakers and listeners could also benefit by looking into other potential non-verbal (pre-beginning) cues such as clicking (see e.g. Ogden, 2013), smacking, swallowing, gestures and gaze with respect to respiratory patterns and their various characteristics.

One of the speakers recorded during these experiments managed to produce a turn with 9 breath groups. Longer speaking turns and their respective inhalation parametres in connection to all of the aforementioned research possibilities would be an interesting set of data to look into. In addition, in this study speakers 3 and 9 provided a very large part of the examined data set. An aspect to note here is that while their speaking turns were dominant in the analysed data, it does not necessarily imply that they were the dominant interlocuters of their conversations. Speaker 3 was asked many questions by other participants who in that sense were leading the course of interaction more. Speaker 9, on the other hand, is indeed a very fast speaker and gave very few possibilities for others to interrupt.
Therefore, the extent to which conversational intentions are realised is very dependent on the other participants. However, this and other possible developments would require a larger corpus.

Each of the described research possibilities might show slightly different results depending on the type of data used – as is known from the different communicative behaviour patterns between cultures (as discussed by Pajupuu, 1995, and Ryabina, 2008), the use of turn-organisatory cues is subject to some variation, depending on cultural norms with respect to aspects such as sensitivity towards silence duration (see e.g. Stivers et al., 2009). Data collected from speakers of different cultures might show that some turn-taking cues are more prevalent than others, and that there are some common combinations of these cues making the intentions of speakers clearer and stronger. The extent to which the cues in turn-organisation need to be intensified might be influenced by the background of the interlocutors.
7 Conclusions

The aim of this thesis was to determine whether respiratory patterns found in spontaneous multiparty conversations could indicate a systematic way of using lung volume amplitudes for turn-organisatory purposes.

The data used for this research consisted of four spontaneous three-party conversations between healthy 22- to 35-year-old native speakers of Estonian, recorded in the Phonetic Laboratory at Stockholm University. Respiratory signal from the conversations was collected using Respiratory Inductance Plethysmography (Watson, 1980). The focus was on uninterrupted speaking turns consisting of multiple breath groups to compare measurements from several consecutive inhalation phases of a turn. In particular, lung volume values at the start and end of each such inhalation phase were of interest, as well as the corresponding amplitudes.

The main research question under investigation was whether inhalation amplitude differs between turn-initial and turn-medial inhalations. Furthermore, if inhalation amplitude proved to show variation, another aim of this thesis was to determine if these differences are reflected by inhalation start and end lung volume values. The data at hand was limited by the small number of recordings and certain conditions applied to the raw data, and, consequently, analysis was based on two inhalation phases: the turn-initial and the first turn-medial inhalation.

Results provide evidence supporting the hypothesis that inhalation amplitude might function as a turn-taking cue. Inhalation amplitude was significantly greater for the inhalation preceding turn onset than for the following inhalation phase in the turn. Furthermore, the increase in amplitude was achieved by ending the inhalations at significantly different lung volume levels: the turn-initial inhalation ended at a higher lung volume value than the first turn-medial inhalation. The inhalation start lung volume levels did not show significant differences between the turn-initial and the first turn-medial inhalation. Results also indicate that the lower inhalation amplitude within turns could be a phrasing or grouping function.

The results thus give reason to continue exploring the possible communicative intentions behind the respiratory patterns speakers use in spontaneous conversations. These patterns could reveal more precise systems and combinations of applying several turn-organizatory devices at relevant dialogue events in order to communicate their intentions and make them more detectable for the other participants.
References


GoPro, Inc. (2013). GoPro HERO 3+ Silver camera.


Register

Figures

Fig. 1. Lung volumes and lung capacities ................................................................. 4
Fig. 2. An example of breathing patterns ................................................................. 6
Fig. 3. Example of recording for one speaker in LabChart ....................................... 13
Fig. 4. Number of turns consisting of multiple and single breath groups .................. 16
Fig. 5. Number of breath groups per speaker left for analysis after filtering raw data .... 17
Fig. 6. Speaker-specific normalised inhalation amplitudes (% VC) ......................... 18
Fig. 7. Normalised inhalation start levels (% VC) according to no. of inhalation ........... 19
Fig. 8. Normalised inhalation end levels (% VC) according to no. of inhalation .......... 20
Fig. 9. Normalised inhalation amplitude (% VC) for inhalations in different locations .. 21
Fig. 10. A comparison of the normalised start and end lung volumes (% VC) for the first two inhalation phases of a turn ................................................................. 22
Fig. 11. A scatterplot of the relationship between normalised inhalation start and end lung volume levels (% VC) ................................................................. 23

Tables

Table 1. Inhalations per speaking turn ................................................................. 17
Table 2. Speaker-specific normalised lung volume values (% VC) ......................... 18
Appendix: Consent form

Name ........................................................................................................................................
E-mail ........................................................................................................................................
Date of birth ................................................................................................................................
Mother tongue ...........................................................................................................................
Other languages  
(Have you lived in another language environment for a longer time?)  
........................................................................................................................................................

Have you suffered from respiratory disorders or difficulties? Yes ☐ No ☐
If yes, then what kind, when? ........................................................................................................

Do you smoke? Have you ever been a smoker? Yes ☐ No ☐
If yes, then when and for how long? ............................................................................................

Have you suffered from speech and/or hearing disorders? Yes ☐ No ☐
If yes, then what kind, when? ........................................................................................................

Information for body mass index calculations:

Weight ..................... (kg)
Height ..................... (cm)

How long have you known your conversation partners? Are you friends, co-workers, family members, ... ?
........................................................................................................................................................
........................................................................................................................................................

I hereby state that my participation in this experiment is voluntary. I know that I can withdraw from participation at any given moment. I allow Stockholm University Phonetics Laboratory record me in audio, video, and record my respiratory signals. I have no information about the exact purpose of the study. I am aware that information gathered in the experiment and on this form is used only for scientific purposes, and my personal information will not be published. Collected data may be used for presentations at scientific conferences and for teaching.

Signature: ............................ Date: ............................................................