Is breathing silence?

Mattias Heldner and Marcin Włodarczak
Department of Linguistics, Stockholm University, Sweden
heldner@ling.su.se, wlodarczak@ling.su.se

Abstract

This paper investigates whether inhalation noises are treated as silences in speech communication. A perception experiment revealed differences in pause detection thresholds for breathing pauses and silent pauses. This in turn indicates that breathing pauses are treated differently by the perceptual system, and could potentially carry a communicative function.

Introduction

Breathing is the primary driving force in speech production, but the respiratory activity in itself can also produce audible noises—both inhalatory and exhalatory. The question we address in this paper is whether such breathing noises are treated as silences in speech communication. The answer is by no means evident. On the one hand, as speakers and listeners we observe that breathing noises may indeed be noticeable. On the other hand, such breathing noises frequently occur in the 'pauses' of speech. We approach this question by investigating whether presence of a breathing noise affects the pause detection threshold in a perception experiment.

Previous research has indicated that the pause or silence detection threshold is located somewhere in the 100–300 ms range. For example, Ruder (1973) estimated the pause detection threshold to be around 95 ms. He also noted that the threshold varies with syntactic complexity and suggested functional differences in within- and between-phrase pauses. However, Ruder's threshold was much lower than the value previously arrived at by Agnello (1963), who set it as high as 190 ms.

Even higher duration thresholds have been found in a conversational setting. According to Walker and Trimboli (1982), the minimal perceptible gap accompanying speaker change in conversation is between 200 and 300 ms. In contrast, in a more recent study Heldner (2011) found that duration thresholds for both gaps and overlaps in dialogue are about 120 ms.

Earlier attempts at estimating pause detection thresholds have instrumentally manipulated pause duration between stretches of speech, resulting in *silent* intervals. This is clearly at odds with the fundamental observation that speakers tend to use pauses to inhale. According to Grosjean and Collins (1979), at normal reading rates (about 200 words per minute) pauses not accompanied by breathing account for only 20% of all pauses, and their frequency drops further as the speech rate increases. While Walker and Trimboli (1982) and Heldner (2011) sourced their experimental material from real conversations, presence of breath was not controlled for in any systematic fashion.

In this study we aim to fill this gap by contrasting detection thresholds for silent and breathing pauses. Notably, we are not so much interested in precise estimation of detection threshold values. Rather, we investigate whether and to what extent presence of inhalation noise affects those thresholds. There are two possible outcomes. If presence of inhalation noise does not affect the pause detection threshold, that would indicate that breathing is treated as silence in speech communication. If, on the other hand, presence of inhalation noise does affect the pause detection threshold, that would indicate that breathing noise is different from and treated as something other than silence.

Why is this question interesting? To us the answer has bearing on the communicative relevance of breathing sounds. Respiratory sounds have long been claimed to be perceptually salient and to play a communicative role in conversation, for instance by cuing speaker changes in conversation (Local & Kelly, 1986; Schegloff, 1996). Indeed, in a recent study (Włodarczak & Heldner, 2016), we have demonstrated that inhalation noises collected with a standard head-mounted microphone can be used to predict speech with accuracy similar to that of kinematic measures of respiratory activity. With this study, we want to establish whether there are perceptual grounds for communicative functions of breathing acoustics.

Method

Five short phrases read by the first author were used as a basis for the stimuli in the perception experiment:

1. *Jag har en överraskning till dig. # Vill du veta vad det är?*
2. *desto tätare svepte vandraren kappan om sig, # och till slut gav nordanvinden upp försöket*
3. *som svätte vatten på publiken, # hästar, lindansare, magiker och clowner*
4. *Hon tog fart och kastade sig ut, # precis som på plakatet*
5. *den som först kunde få vandraren att ta av sig kappan, # den skulle anses vara starkare än den andra.*

Segments of breath or silence of varying durations were then inserted into the carrier phrases at the location marked by #.

Breath segments were constructed by removing the middle part of a long (> 500 ms) pre-speech inhalation from the same recording. Durations of the breath segments used in the experiment were: 50, 75, 100, 125, 150, 175, 200 and 300 ms. The acoustic intensity of the sound segments varied from 33.11 to 34.96 dB (mean = 34.75 dB, sd = 0.56 dB).

Pauses of the same durations were cut from a silent portion of the recording. The breath and silence segments were subsequently spliced into the carrier phrases, giving 80 stimuli.

The perception experiment was implemented as a Praat Multiple Forced Choice Experiment (Boersma & Weenink, 2016). All stimuli were presented once only in random order. The sounds were played through high-quality headphones (Beyerdynamic DT 770 Pro) attached to a computer in a sound-treated studio in the Phonetics Lab, Stockholm University. The loudness level was set so that the inhalations were clearly audible.

The participants were 14 employees or students at the Department of Linguistics, Stockholm University (7 males, 7 females). There were instructed to listen to the stimuli and decide whether the speaker paused at any point during the utterance. The instructions were presented on the screen. The experiment was preceded by a short pre-test comprising of 42 stimuli, which the participants were allowed to stop once they felt they have understood the task. All participants were native or highly proficient speakers of Swedish. The experiment took approximately 15 minutes. The subjects were rewarded with a cinema ticket for their participation.

We used binary logistic regression to calculate the pause detection thresholds for the individual participants as well as for the group of participants. We followed the process of fitting a logistic regression model outlined in Field (2013). The individual and group models all included *pause duration* (i.e. interval between end of preceding speech to onset of following speech), *condition* (breath or silence), and the interaction between the two as predictors of the participants' answers.

Similar to previous research, we used the level corresponding to 0.5 probability of perceiving a pause as the detection threshold (e.g. Gescheider, 1997, pp. 116-117). Needless to say, we calculated separate thresholds for the *breath* and the *silence* conditions.

Results

In the first analysis step, we calculated pause detection thresholds for the 14 individual participants (see Table 1). Four participants showed negative and hence unrealistic detection threshold values and were excluded from further analysis.

Table 1. Individual detection thresholds in silence and breath conditions. Negative detection thresholds are marked in bold.

	Detection Thresholds	
Participant	Silence	Breath
P1	122	150
P2	112	94
P3	74	161
P4	58	108
P5	10	**-67**
P6	98	190
P7	81	134
P8	53	**-34**
P9	92	186
P10	**-43**	109
P11	86	100
P12	**-974**	**-1434**
P13	149	229
P14	90	135

Subsequently, we fitted a binary logistic regression model for the remaining group of ten participants. We built up the model one predictor at a time to ascertain whether the ‘new’ model improved the fit compared to the previous model. This process revealed that both *pause duration*, *condition* and the interaction between the two made significant *($p < 0.05$)* contributions to the model fit according to the change in log-likelihood. The corresponding group detection thresholds were 100 ms for the silent condition and 142 ms for the breath condition. The fitted logistic functions used to determine the detection thresholds are shown in Figure 1.

The results thus showed that duration significantly influenced the probability that a subject would perceive a pause. Furthermore, the probability was also influenced by presence of inhalation noise. More importantly, the two main effects were not independent in that the effect of duration was different in the two conditions. In other words, the pause needed to be longer in the breath condition than in the silence condition in order to be perceived. The same qualitative pattern of higher thresholds in the breath condition was also observed for nine out of ten participants (cf. Table 1).

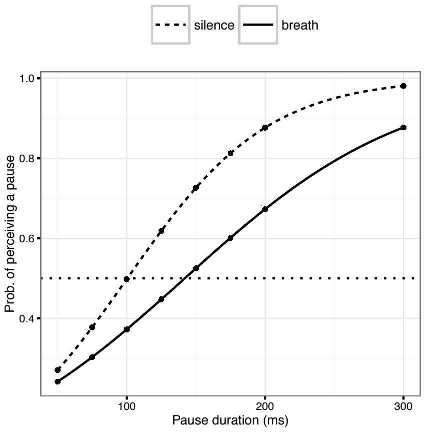

Figure 1. Estimated probabilities of pause detection with (breath) and without (silence) breathing noises given the duration of the pause (from a logistic regression analysis). The detection threshold is the duration at which the estimated probability equals 0.5 (dotted line).

Discussion

With respect to the question in the title, our results indicate that breathing pauses are perceptually different from silent pauses. Breathing is thus *not* silence.

Importantly, given that in spontaneous speech, most pauses are accompanied by breathing, our study provides a more ecologically valid estimation of pause thresholds than those reported previously (e.g. Agnello, 1963; Ruder, 1973).

Furthermore, our results have implications for what information is relevant for turn-taking in conversation. While this experiment was not concerned with functional distinctions, by demonstrating perceptual differences, the study highlights the possibility that breathing pauses might carry a communicative function. For example, since breathing pauses have higher detection thresholds, they could be used as a turn-keeping device. Specifically, breathing pauses could be a way of occupying the transition relevance place/space to lower the likelihood of pause interruption by another speaker. This hypothesis is plausible given that inhalation duration tends to be shorter in turn internal pauses (Rochet-Capellan & Fuchs, 2014) leading to higher flow rate and consequently to greater acoustic intensity of breath (Włodarczak & Heldner, 2016).

Consequently, the results provide the perceptual basis for the claims of communicative relevance of respiration. As such, they bring us closer to understanding the manifold functions of breathing in conversation.

Acknowledgements

This work was funded by the Swedish Research Council (VR) project 2014-1072 *Andning i samtal (Breathing in conversation)*.

References

Agnello, J. (1963). *A study of intra- and inter-phrasal pauses and their relationship to the rate of speech.* (Unpublished dissertation), Ohio State University.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.18) [Computer program]. Retrieved from http://www.praat.org/

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (fourth ed.). Los Angeles: Sage.

Gescheider, G. A. (1997). The classical psychophysical methods *Psychophysics: The Fundamentals* (Third ed., pp. 45-72). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica, 36*, 98-114.

Heldner, M. (2011). Detection thresholds for gaps, overlaps and no-gap-no-overlaps. *Journal of the Acoustical Society of America, 130*(1), 508-513. doi: 10.1121/1.3598457

Local, J., & Kelly, J. (1986). Projection and ‘silences’: Notes on phonetic and conversational structure. *Human studies, 9*, 185-204.

Rochet-Capellan, A., & Fuchs, S. (2014). Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B-Biological Sciences, 369*(1658), 1–10. doi: 10.1098/rstb.2013.0399

Ruder, K. F. (1973). Duration of silent interval as a perceptual cue of speech pauses. *Perceptual and Motor Skills, 36*, 47-57.

Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 52-133). Cambridge: Cambridge University Press.

Walker, M. B., & Trimboli, C. (1982). Smooth transitions in conversational interactions. *The Journal of Social Psychology, 117*, 305-306.

Włodarczak, M., & Heldner, M. (2016). Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking *Proceedings Interspeech 2016*. San Francisco, USA: ISCA.