Speech Communication 88 (2017) 127–136

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Effects of noise on arousal in a speech communication setting☆

Nemanja Cvijanović[a,b,*], Patrick Kechichian[a], Kees Janse[a], Armin Kohlrausch[a,b]

[a] Philips Research Laboratories, Eindhoven, The Netherlands
[b] Human-Technology Interaction, Eindhoven University of Technology, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 18 July 2016
Revised 15 December 2016
Accepted 2 February 2017
Available online 3 February 2017

Keywords:
Arousal
Heart rate
Skin conductance
Speech communication
Noise
Communication effort

ABSTRACT

Speech communication in natural environments is often impaired by varying levels of ambient noise. Such noise can reduce speech intelligibility and make conversations more effortful, consequently causing an increase in arousal, frustration or stress in the partaking speakers. This contribution investigates the effects of background noise on arousal in a speech communication setting using collaborative tasks and examines the measurability of such detrimental effects through physiological signals – heart rate variability and skin conductance. We focus on the differences in responses between the various noise conditions aiming to establish the sensitivity of the employed physiological measures. Furthermore, self-reported mental effort scores are used to assess the dependency of subjective mental effort requirements on background noise for the examined communication setting. Our results indicate that while mental effort scores show a significant positive relationship with background noise level, skin conductance and heart rate variability features, which are commonly employed for arousal state detection, may be inappropriate and lack sensitivity to distinguish communication settings differing solely by the level of background noise.

1. Introduction

In real world communication scenarios, the presence of ambient noise is inevitable and has detrimental effects on speech communication. These can range from a decline in speech intelligibility, resulting in an increase in the required listening effort, to stress and increases in arousal and annoyance (Sarampalis et al., 2009; Stansfeld and Matheson, 2003; Basner et al., 2014). Being a well-known problem in speech communication systems, numerous methods which suppress ambient noise without degrading the speech signal have been developed in the past (Loizou, 2007).

The performance of noise suppression methods and their benefits are usually evaluated using objective metrics like the segmental signal-to-noise-ratio (SNR) gain (Hansen and Pellom, 1998) or the perceptual evaluation of speech quality (PESQ) (ITU-T recommendation, 2001), among others (see Hu and Loizou, 2008 for a more detailed review on objective evaluation of speech enhancement methods). However, while some of the objective metrics like PESQ show high correlations with subjective measures of speech quality obtained during listening tests (Hu and Loizou, 2008), they often convey little information about the user's personal experience and how it is affected by background noise level, e.g., when the system does not respond as expected. The present study examined this in more detail, by emulating a realistic communication setting under varying background noise levels with the aim to manipulate the so-called communication effort required for successful communication. The effects of the background noise were then assessed through the user's arousal states, where arousal is defined as a state of alertness, readiness and responsiveness to stimuli (Martin, 2015).

To define communication effort and its connection to arousal, we can start with the more general concept of mental effort, which has been recently defined as *the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task* (Pichora-Fuller et al., 2016). Listening effort is then defined as *a specific form of mental effort that occurs when a task involves attending and understanding an auditory message* (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). The connection to communication effort can then be established by expanding this definition to *a specific form of mental effort that occurs when a task involves attending and understanding an auditory message as well as the formulation and production of a meaningful response*. An increase in background noise is expected to result in an increase in listening effort, as shown in numerous studies in the literature (Sarampalis et al., 2009; Zekveld et al., 2010; Pals et al., 2015), which then results in

☆ This project has been funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000 and is part of the INSPIRE Initial Training Network.

* Corresponding author.
E-mail addresses: nbcvijanovic@gmail.com (N. Cvijanović), patrick.kechichian@philips.com (P. Kechichian), kees.janse@philips.com (K. Janse), armin.kohlrausch@philips.com (A. Kohlrausch).

http://dx.doi.org/10.1016/j.specom.2017.02.001
0167-6393/

an increase in communication and mental effort, according to the presented definitions.

A similar positive relationship also exists between mental effort and arousal (Kahneman, 1973) and has been used in numerous task-based arousal detection and monitoring studies to increase arousal through the increase of the task difficulty (required mental effort), e.g., Taelman et al., 2011; Wood et al., 2002; Sun et al., 2012; Alexandratos et al., 2014; Zhai and Barreto, 2006. An increase in background noise during communication is thus expected to lead to an increase in arousal. This assumption is further supported by recent developments in the assessment of listening effort.

In psychophysiological studies, arousal is usually assessed through the recording and analysis of a variety of physiological signals which reflect changes in the activity of the autonomic nervous system (ANS), or more specifically, in the activity of two of its branches – the sympathetic nervous system (SNS) and the parasympathetic nervous system (PSNS) (Blascovich and Kelsey, 1990). Common approaches include the use of electrocardiograms (ECG) and photoplethymograms (PPG) and subsequent heart rate and heart rate variability (HRV) analysis (Melillo et al., 2013; Sun et al., 2012; Healey and Picard, 2005; Alexandratos et al., 2014; Wood et al., 2002; Taelman et al., 2011; Vrijkotte et al., 2000; Frank et al., 2013), the capture of electro dermal activity (EDA) signals and the analysis of corresponding changes in skin conductance (Healey and Picard, 2005; Zhai and Barreto, 2006; Sun et al., 2012; Ogorevc et al., 2011; Alexandratos et al., 2014), pupillometry (Zhai and Barreto, 2006; Ren et al., 2014), detection of changes in skin temperature (Zhai and Barreto, 2006; Ogorevc et al., 2011; Frank et al., 2013) and the detection of changes in blood pressure(Vrijkotte et al., 2000; Ogorevc et al., 2011), among others. These physiological measures have recently also been used for measuring listening effort, e.g. skin conductance in (Mackersie and Cones, 2011; Mackersie et al., 2015; Seeman and Sims, 2015) and HRV in Mackersie et al. (2015); Seeman and Sims (2015); Dorman et al. (2012), with the rationale that changes in listening effort affect the activity of the ANS and should be measurable through physiological signals (McGarrigle et al., 2014).

In this contribution, we aim to detect changes in arousal states elicited by noise in a realistic communication setting, by recording and analyzing electrocardiogram (ECG) and EDA signals from participants solving collaborative communication tasks together under different noise conditions. The physiological signal choice is motivated by recent technological developments and the existence of numerous wearable sensors which provide heart rate (HR) and skin conductance (SC) signals in real time (e.g., Emp, 2016). Such sensors could be utilized in real life communication scenarios to measure the effects of background noise and potentially adapt the characteristics of noise suppression algorithms on-line to aid the user. In addition to physiological measures of arousal, a subjective self-report measure of mental effort is used to assess changes in effort required for task completion across different noise conditions. Finally, this work also aims to investigate the sensitivity of a selection of the most commonly used HRV and EDA features used for detecting changes in arousal states.

While numerous studies in the literature investigate the differences in arousal between substantially different (e.g., activity vs. inactivity) baseline and arousing conditions (Taelman et al., 2011; Sun et al., 2012; Alexandratos et al., 2014; Zhai and Barreto, 2006) or differences between substantially different types of tasks (e.g., mental vs. physical) (Taelman et al., 2011; Vrijkotte et al., 2000; Ogorevc et al., 2011), our experiment examines the changes in arousal across conditions that differ solely by the level of background noise. The employed collaborative communication tasks are used to stimulate communication between participants and their difficulty level remains similar throughout all conditions. This ensures a comparable standard arousal level and participant engagement across all noise conditions, making the level of background noise the main differentiating factor.

To the best of the authors knowledge, neither the effects of background noise on arousal nor the sensitivity of commonly used physiological measures have been previously investigated in a realistic setting involving collaborative communication tasks. Hereby, the presented experiment has two hypotheses. The first hypothesis is used to validate our experimental design and states that an increase in background noise leads to an increase in communication effort and consequently in mental effort. This hypothesis is analyzed through the subjective mental effort measure. The second and main hypothesis is that an increase in background noise leads to a measurable increase in participant arousal and it is assessed through the analysis of the captured physiological data.

2. Experimental design and methodology

The experimental setup for our study relied on speech communication under different noise conditions between two participants. These were seated in separate rooms and communicated with each other using headphones and a microphone placed on the table in front of them. This setup was designed to emulate a hands-free communication scenario, e.g., a Skype® conversation. During the experiment, the participants were given collaborative tasks to be solved together under three different background noise conditions. After each condition, the participants filled out mental effort self-evaluation scales. They were given time to relax in between conditions and before the start of the experiment. Furthermore, participants completed a practice run to get used to the tasks and the laboratory environment prior to the experiment. Throughout each session, ECG and EDA signals from each of the participants were recorded and later used to detect changes in arousal states. The study procedure was approved by the ethics and review boards of Philips Research Laboratories Eindhoven and the Eindhoven University of Technology. The timeline of a typical experimental session is shown in Table 1 and various aspects of it are explained in the following.

2.1. Experimental conditions

Throughout the experiment, three different background noise conditions were used – silent, 6 dB and −6 dB SNR. In the latter two, noise was added in the communication channel at a level required to reach the corresponding SNR of the noisy speech signal output. Hereby, a looped version of the 30 s non-stationary cafeteria noise signal specified in UMTS (2012) was used.

The noise level in the noisy conditions was adjusted on each side of the communication channel based on the speech signal level captured by the corresponding (participant's) microphone. Before the start of the experiment, participants were asked to briefly communicate in the −6 dB SNR noise condition and set a comfortable headphone output level, to which they could be exposed to for a prolonged period of time. The resulting headphone output level, L_t, was then used throughout the experiment as an upper threshold for the noise level to avoid excessive discomfort and annoyance of the participants, especially in the −6 dB SNR condition. During noisy conditions, the transmitted noisy speech signal $y(k)$ is given as $y(k) = g_s \cdot s(k) + g_n \cdot n(k)$, where $s(k)$ and $n(k)$ are the captured speech signal and the noise signal, respectively. The corresponding gains g_s and g_n are set to adjust the resulting SNR of the signal played back over the headphones given by

$$\mathrm{SNR}_{\mathrm{out}} = 10 \cdot \log_{10}\left(\frac{g_s \cdot \sigma_s^2}{g_n \cdot \sigma_n^2}\right), \quad (1)$$