Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 56 (2014) 35–48

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Ascribing emotions depending on pause length in native and foreign language speech

Eszter Tisljár-Szabó *, Csaba Pléh [1]

*Budapest University of Technology and Economics, Department of Cognitive Science, Egry József u. 1, H-1111 Budapest, Hungary*

Received 31 January 2012; received in revised form 8 July 2013; accepted 29 July 2013
Available online 8 August 2013

**Abstract**

Although the relationship between emotions and speech is well documented, little is known about the role of speech pauses in emotion expression and emotion recognition. The present study investigated how speech pause length influences how listeners ascribe emotional states to the speaker. Emotionally neutral Hungarian speech samples were taken, and speech pauses were systematically manipulated to create five variants of all passages. Hungarian and Austrian participants rated the emotionality of these passages by indicating on a 1–6 point scale how angry, sad, disgusted, happy, surprised, scared, positive, and heated the speaker could have been. The data reveal that the length of silent pauses influences listeners in attributing emotional states to the speaker. Our findings argue that pauses play a relevant role in ascribing emotions and that this phenomenon might be partly independent of language.


*Keywords:* Emotion ascribing; Cross-linguistic; Speech pauses; Silent pause duration; Foreign language

## 1. Introduction

The goal of vocal emotion research is to understand the relationship between emotions and certain speech parameters. Most of these studies are encoding or decoding studies (for reviews see Juslin and Scherer, 2005; Scherer, 2003). In the case of the encoding studies, the acoustic pattern of expressions uttered in different emotional states is analysed, while in the case of the decoding studies, the focus is on the recognition of emotions. Studies in the area of decoding examine whether listeners are able to recognise emotions in speakers' speech and if so, which parameters help listeners with this activity.

Although the relationship between emotions and speech is well documented (Juslin and Laukka, 2001; Johnson et al., 1986; Bänziger and Scherer, 2005; Banse and Scherer, 1996; Schröder, 2003), few studies have examined the role of speech pauses in emotion expression and emotion recognition. In the present study, the main question concerned how listeners' emotion judgments are influenced by the length of the speech pauses. In addition, a second aim was to compare two different nations (Hungary and Austria) on the same set of stimuli to address questions about sample differences in emotion decoding according to speech pause length. To examine this concept, we conducted two experiments on the same speech material: one in a native context and one in a foreign language context.

### 1.1. Emotions and speech

The combination of coding and decoding studies is used as another method to examine the effect of change in certain speech parameters on listeners' emotion judgments (for an overview see (Scherer, 2003). For example, with

* Corresponding author. Present address: University of Debrecen, Medical and Health Science Center, Department of Behavioural Sciences, Nagyerdei krt. 98, P.O. Box 45, H-4032 Debrecen, Hungary. Tel.: +36 52 411 717x56531; fax: +36 52 255 723.

*E-mail addresses:* jszaboeszter@gmail.com (E. Tisljár-Szabó), pleh.csaba@ektf.hu (C. Pléh).

[1] Present address: Eszterházy Károly College, Eszterházy tér 1, H-3300 Eger, Hungary.

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.07.009

the help of computer programs, acoustic cues in neutral or meaningless sample utterances are systematically manipulated, and researchers examine how this phenomenon influences listeners in terms of the type and strength of the emotions that they ascribe to the speaker. There are many parameters to manipulate, including fundamental frequency (F0), range, intensity, emphasis, and temporal cues (Ladd et al., 1985; Bergmann et al., 1988; Breitenstein et al., 2001).

For example, Bergmann et al. (1988) manipulated F0 range, jitter, intensity, intonation contour, and duration in expressions produced by actors. The listeners' task was to rate the emotionality of the speaker with the help of a rating scale with labels for emotion and attitude categories. According to their main results, if F0 range increased, the speakers were rated as more emphatic, angrier, and more aroused, while if F0 range decreased, speakers were rated as sadder. If the intensity of the sentence increased, ratings on reproachful, angry, and empathic scales were higher. The systematic modification of accent-high had the most effect on scales that measured cognitive state or attitude. The higher the accent was, the more contradictory, the more reproachful, and the more empathic the speaker was rated. In the study, the length of the accent was also modified by lengthening its vowel. According to the results, a short vowel length leads listeners to rate speakers as happier, while a long vowel length leads listeners to rate the speakers as sadder.

In other studies, both neutral and emotionally loaded sentences are modified (Bergmann et al., 1988; Breitenstein et al., 2001). Breitenstein et al. (2001) used neutral and emotionally loaded sentences belonging to four emotion categories from a previous study (Breitenstein et al., 1996). Stimulus material was compiled by systematic variation of the F0 variation and speech rate of the sentences in six ways for both parameters. The listeners' task was to rate the emotional tone of voice of each sentence by choosing one of the emotion labels. The results showed that despite the manipulation, subjects most frequently chose the emotion label that corresponded to the emotion category in which the speaker had originally made the utterance. Recognition of happy and sad sentences was most markedly affected by increased speech rate. When the speed of the originally happy sentence was increased, the rate of neutral, frightened, and angry answers increased. When the speed of the originally sad sentences was increased, again, the rate of neutral and frightened answers increased. In most of the sentences, independent of the emotional category, if the utterance became slower, then the frequency of sad ratings increased, while the frequency of frightened, angry, and neutral ratings decreased.

The results of the above mentioned studies show us that modifications in even one speech parameter can change listeners' opinions about the speaker; moreover, listeners are able to perceive changes in temporal parameters of speech, and accordingly, they ascribe different emotional states to the speaker.

### 1.2. Relationship between speech pauses and emotions

Only a few studies have investigated the role of speech pauses, most of which have studied the relationship between pauses and the cognitive activity needed for speech. According to Goldman-Eisler (1958, 1968) and Rochester (1973), an increase in speech disfluencies and pauses is related to increased cognitive activity. Thus, pauses are more likely to occur before a word that is harder to access in the mental lexicon, or before a syntactically harder expression where the verbal process requires greater cognitive effort.

The relationship between speech pauses and emotions is the research topic of several approaches. Some researchers (Eldred and Price, 1958; Kasl and Mahl, 1965; Mahl, 1956; Pope et al., 1970) have investigated the relationship between speech errors and emotions, and they count silent pauses as one type of error. Other so-called vocal emotion expression studies (Fairbanks and Hoaglin, 1941; Jovicic et al., 2004; Szabó, 2008) investigate the presence of different emotions in speech or voice, and speech pauses appear as part of the speech parameters. A third approach is linguistic conversation analysis, which has also been conducted in studies investigating the role of speech pauses during emotionally loaded situations (Deppermann and Lucius-Hoene, 2005).

The early studies that investigated the relationship between speech disfluencies (or speech errors) and anxiety found that in an anxious state, the number and length of pauses increase (Eldred and Price, 1958; Kasl and Mahl, 1965; Mahl, 1956; Pope et al., 1970). Later, Hofmann et al. (1997), as well as Laukka et al. (2008), analysed the temporal characteristics of the speech of social phobic patients with public speaking anxiety. According to Hofmanns' study (Hofmann et al., 1997), social phobics produce longer speech pauses, pause more frequently, and spend more time pausing than do controls when giving a speech. Laukka et al. (2008) compared speech samples from social phobic patients before and after pharmacological treatment. The patients' task was to give a speech about a vacation or travel experience before an audience of six to eight people. According to the results, the rate of silent pauses decreased in the speech samples of people who responded to the treatment, while in the case of non-responders, the pause ratio increased. In connection with speech tempo, the researcher did not find any significant differences.

Deppermann and Lucius-Hoene (2005) analysed trauma narratives from TV shows using conversation analysis. They found that there were more pauses in speakers' speech while they were telling sad stories. Speakers even took longer pauses within syntactical units. In Fairbanks and Hoaglin (1941) study, a 27-word-long passage was spoken by six male actors, simulating five emotional states (contempt, anger, fear, grief, and indifference). The observation of durational features was the main focus of that study. According to the results, fear, indifference, and