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

anger had the fastest tempo, followed by grief and contempt. Fear, indifference, and anger were characterised by a rapid rate, a short duration of phonations and short pauses with much variability, while in the cases of contempt and grief, a slower speech rate was measured. Contempt and grief, however, differed in terms of the cause of this slower rate. In the case of contempt, the slower speech rate was caused by prolongation of both phonations and pauses, and the rate of pauses was approximately the same as the three other emotions. However, the slow rate of grief was caused almost entirely by prolongation of pauses; thus, the pause ratio significantly differed from that of the other emotions. In the case of grief, the rate of pauses was greater than 50% in the speech of three of the actors.

Jovicic et al. (2004) established a Serbian emotional speech database. They asked actors to simulate emotional states while uttering words, sentences, and a passage. Speech duration, pause duration, and speech/pause ratio were studied and measured. Researchers found that pause duration was more discriminative between emotions than speech duration and that the speech/pause ratio was a good indicator for speech rhythm and differentiated well between angry, happy, frightened, sad, and neutral emotional states. In Szabó's research (Szabó, 2008), sad and happy emotional states were induced by using music and autobiographical recall tasks, and half-minute long speech samples were analysed. The results showed that in a sad emotional state, the average length of silent pauses and the rate of pauses to the whole speech increased compared to a happy emotional state.

The results of the above mentioned studies indicate that length and rate of speech pauses can change in different emotional states. However, most of these studies are connected by the emotional states of anxiety or fear, and little is known about other states such as happiness, sadness or anger.

1.3. Previous studies on vocal emotion recognition in foreign languages

Many studies have investigated emotion recognition in different cultures, most of which have focused on whether subjects belonging to different cultures can recognise expressions of emotion on the faces of others (Ekman et al., 1969; Izard, 1971, 1994; Ekman and Strom, 1969; Ekman and Friesen, 1971; Biehl and et al., 1997). These studies demonstrated that although emotion expression is regulated by display rules, cultural norms, and individual habits, individuals are able to accurately recognise emotional displays of the face independent of different cultural backgrounds. Thus, these studies argue that emotional display involves universal principles.

There are fewer studies that have investigated how individuals from different linguistic and cultural backgrounds recognise one another's emotions from vocal expressions (Beier and Zautra, 1972; Vanbezooijen et al., 1983; Mccluskey and Albas, 1981; Mccluskey et al., 1975; Thompson and Balkwill, 2006; Scherer et al., 2001). In these studies, generally the emotionality of sample utterances by lay people or professional actors is judged, and researchers compare how the accuracy in recognising emotions differs when the speaker and listener share cultural and linguistic background and when they do not.

In Beier and Zautra's study (Beier and Zautra, 1972), American, Japanese and Polish subjects rated the emotionality of English sentences that were neutral based on content. Their results showed that when the judged expressions were short, Americans were better at emotion recognition than Japanese and Polish subjects. At the same time, by increasing the length of the expressions, the differences disappeared. Vanbezooijen et al. (1983) asked Dutch, Japanese, and Taiwanese subjects to rate the emotionality of sentences simulated by Dutch actors. The results were similar to those of the previous study: native speaking Dutch subjects performed significantly better than the other two groups, but Japanese and Taiwanese participants were also able to recognise the majority of emotions with above-chance accuracy, and the pattern of confusion was similar across judgement groups.

Later studies used so-called pseudo-utterances to filter out the effect of sentence meaning. The pseudo-utterances were made by connecting meaningless syllables from different languages in such a way that utterances that sounded like speech, but were semantically meaningless, were created. In a study by Scherer et al. (2001), professional German actors simulated five emotions by uttering pseudo-sentences. Subjects from nine countries participated in the experiment, and their task was to judge the emotional content of the utterances on five emotion scales ranging from 0 to 6. Their results showed that German participants performed significantly better than the other language groups; thus, accuracy was highest when the speaker's and listener's language was the same. Further results showed that participants, whose language was more linguistically different from German, performed less accurately than those from a highly similar language. Thus, German subjects performed the best, followed by the Swiss-French, English, Dutch, American, Italian, French, and Spanish groups. The worst accuracy level was produced by the Indonesian subjects, but even this group performed with better than chance accuracy. Based on these findings, the authors formulated the "language distance hypothesis", which says that language similarity plays an important role in emotion decoding in a foreign language.

In another study conducted by Pell et al. (2009), pseudo-utterances from native speakers of English, German, Hindi, and Arabic were rated by subjects from the same languages. Emotion recognition and acoustic patterns were analysed. The results showed that the mean, variance in fundamental frequency, and speech rate were responsible for 70–80% of the variance in the acoustic data across emotion types. The authors argue that these parameters determine the vocal expressions of basic emotions on a large scale, which might be a universal cue.

Based on previous studies (Vanbezooijen et al., 1983; Thompson and Balkwill, 2006; Scherer et al., 2001; Albas et al., 1976; Pell and Skorup, 2008; Fónagy and Magdics, 1963), we argue that listeners can accurately detect and categorise emotional states from speakers with a different language and cultural background. Processes of vocal emotion recognition involve universal principles; thus, the phenomenon is at least partly universal. However, some studies (Beier and Zautra, 1972; Vanbezooijen et al., 1983; Scherer et al., 2001) have reported an in-group advantage, as vocal emotions that are simulated by speakers of the same culture and language are more accurately identified when compared to speakers of a different culture and language. Therefore, beyond universality, social aspects and language-specific prosodic features are also important in communicating emotions (Thompson and Balkwill, 2006; Beaupre and Hess, 2005; Ekman et al., 1987; Matsumoto, 1993; Elfenbein et al., 2007). This phenomenon is described by the "language distance hypothesis" of Scherer and collaborators (Scherer et al., 2001) and by the "cultural proximity" hypothesis of Elfenbein and Ambady (2003) as well. According to this last hypothesis, people who are more similar in their cultural background can more accurately decode one another's emotions than people whose cultures are different.

1.4. Cross-section of intercultural and inference studies

The research of Breitenstein et al. (2001) is a good example of the cross-section of intercultural and inference studies. As mentioned previously, in Breitenstein and collaborators' experiment, the speech rate and variance of the fundamental frequency of emotional sentences produced by a German actress were systematically manipulated. The emotionality of utterances was judged not only by German raters, but also by American raters. According to the results, the judgments of the two groups did not greatly differ, as despite the manipulation, subjects most frequently chose the emotion label that corresponded to the emotion category in which the speaker had originally spoken the utterance. The only exception was for the happy sentence, which American participants most frequently placed in the neutral category. Further analysis of the data showed that American raters were more influenced by tempo manipulations than Germans. For example, in the case of the happy sentence, German participants more frequently chose the happy category at faster rates, while American raters chose the angry or frightened answers more frequently. If the rate of angry sentences decreased, Americans were less accurate than the Germans and often responded with the neutral category. For frightened and neutral sentences at a slower rate, Americans more often answered with the sad category than did Germans. Manipulation in F0 variation influenced German and American subjects in the same way, and there were only small differences between the two groups.

In a study by Burkhardt et al. (2006), French, German, Greek, and Turkish sentences were systematically manipulated with respect to pitch range, duration, and jitter. In a perception experiment, French, German, Greek, and Turkish subjects listened to the original and the manipulated utterances, each in the subjects' native language, and rated how appropriate the manner of speaking was for the meaning of the sentence (e.g., happy, disgusted, etc.). The results showed that regardless of the language, listeners interpreted the emotional load of the phrases as intended, with the exception of the bored phrases. As expected, phrases with the original prosody were generally judged as more appropriate than those that were manipulated. Furthermore, negative emotions were found to be better displayed by the speech synthesizer than the positive ones. As there were language differences, regarding what kind of pitch range, duration, or jitter manipulation had an effect on the emotion judgements, the authors argue that emotion simulation cannot work the same way in all languages.

1.5. The current study

The present study differs from previous studies in many respects. The so-called inference studies usually manipulate sentences that are uttered by actors (Ladd et al., 1985; Bergmann et al., 1988; Breitenstein et al., 2001; Scherer et al., 1984; Cahn, 1990; Carlson, 1992; Burkhardt et al., 2000). The advantage of using sentences uttered by actors is that it produces well controlled and high quality speech samples with intensive emotions, and changes can be well observed and measured. However, the disadvantage of this method is that these speech samples can differ in numerous ways (e.g. pausing, accentuation, speech errors) from those that occur in real life (see Juslin and Scherer, 2005; Scherer, 2003). Therefore, in our experiment we used natural speech samples, namely, samples from interviews from a Hungarian speech database (Gósy, 2008) and from an experiment (Szabó, 2008) in which speakers were volunteers and talked about themselves for a few minutes. The speech material sounded natural and lifelike compared to emotional expressions by actors. We selected one-minute long monologic sections to study speech pauses.

Several studies (Bergmann et al., 1988; Breitenstein et al., 2001; Burkhardt and Sendlmeier, 2000) have investigated the relationship between tempo manipulation and emotion ascribing, as well as modified articulation rate and speech rate[2] together. In these studies, the articulation rate and possible pauses between sentence units, and thus speech rate, were both modified by lengthening or shortening the sentence. However, there are only few studies

[2] Speech tempo is defined as the number of production units (usually phonemes or syllables) per unit of time (usually seconds or minutes), where total speech time includes pauses; while articulation rate is defined as the number of production units per unit of time after subtracting pauses. (see 52. Crystal, T.H. and A.S. House, *Articulation Rate and the Duration of Syllables and Stress Groups in Connected Speech.* Journal of the Acoustical Society of America, 88 (1), 1990, p. 101–112.)

which examined the role of speech pauses in emotion expression (for an example see Juslin and Laukka, 2001), thus little is known about this topic. For investigating this phenomenon, utterances with naturally occurring pauses are needed. The present study used this kind of material and manipulated the pauses of them. In contrast to previous studies (e.g. Breitenstein et al., 2001), speech tempo was modified only by manipulating silent pauses and by leaving the articulation rate fixed.

Thus, in this study, we investigated how listeners' attributions of the emotional state of speakers are affected by the systematic manipulation of the length of silent pauses in originally natural speech samples. In Experiment I, pauses in Hungarian speech samples were modified in four different ways, and variants were rated by listeners. Because, to our knowledge, there is no work that has investigated the relationship between speech pauses and emotions by using manipulated speech samples, we mostly based our hypotheses on the results of previous encoding and decoding studies (Juslin and Scherer, 2005; Breitenstein et al., 2001; Fairbanks and Hoaglin, 1941; Szabó, 2008; Deppermann and Lucius-Hoene, 2005; Hofmann et al., 1997; Laukka et al., 2008). According to our hypothesis, in the case of longer speech pauses, speakers would be rated as sadder, less happy, less angry, more scared, more disgusted, more surprised, less positive, and less heated.

In Experiment II, the same Hungarian experimental speech material was listened to and judged by Austrian subjects with a German mother tongue. Although Hungary and Austria are neighbouring countries, the German and Hungarian languages belong to different language families, and are very different concerning the vocables or the grammar. Thus for German speaking Austrians, Hungarian sentences were meaningless. By using this kind of stimuli, we had the possibility to investigate whether the role of the pauses in emotion ascribing is a language independent phenomenon or not. Based on previous intercultural vocal emotion recognition studies (Vanbezooijen et al., 1983; Thompson and Balkwill, 2006; Scherer et al., 2001; Pell and Skorup, 2008) and the cultural proximity hypothesis (Elfenbein and Ambady, 2003), we assumed similarities in the relationship between speech pauses and emotions in Hungarian and German. Thus, we hypothesised that if native German speaking subjects listen to the Hungarian speech samples, pause manipulation would have a similar effect on trends in emotion ratings, as in the case of the Hungarians. Thus, in the case of pause elongation, speakers would be rated as sadder, less happy, less angry, more scared, more disgusted, more surprised, less positive, and less heated.

2. Material and methods

2.1. Experimental material

In the experiment, we used five emotionally neutral speech samples that originated from the Hungarian Speech Database (Gósy, 2008) and a previous study we had conducted (Szabó, 2008). On the recordings, five female volunteers talked about their schools, jobs or an ordinary day in their lives. The speech samples were approximately one minute long (49.1–64.4 s, average: 55.8 s). The content was emotionally neutral and did not contain emotional words (such as "*it was a very happy part of my life*") or laughter. The speech samples were all monologues and were extracted from longer interviews. Here we give an example where speaker *A* is talking about an ordinary day of her. Already present speech pauses are marked by <***pause***>.

"*Reggel általában olyan 7 óra felé kelek fel* **<pause>** *utána rögtön az az első, hogy elkészítem a reggelit* **<pause>** *tehát az a legfontosabb, a reggel fénypontja,* **<pause>** *és arra szánok a legtöbb időt is, tehát legalább fél óra kell, hogy én reggelizzek, és utána jöhet a többi készülődés, rendet rakás satöbbi.* **<pause>** *De ugye a nap nagy részét a suliban töltöm,* **<pause>** *az órák mellett gyakran el szoktam menni könyvtárba* **<pause>** *kiegészíteni a* **<pause>** *tanulmányaimat...* **<pause>** *Öööö,* **<pause>** *illetve* **<pause>** *Öööö* **<pause>** *Igen* **<pause>** *És az iskola után pedig* **<pause>** *bevásárolni igen gyakran szoktam menni..* **<pause>**. *Utána,* **<pause>** *Öööö* **<pause>** *miután hazaérkeztem,* **<pause>** *megfőzök másnapra,* **<pause>** *de ez nem mindig így van, tehát hetente kétszer* **<pause>** *szokott ez lenni. Aztán hogy ha kell valamit készülnöm másnapra, azt is megteszem,* **<pause>** *és utána pedig vagy a párommal elmegyünk valahová* **<pause>,** *vagy éppen táncolni, amit említettem, vagy moziba* **<pause>**, *vagy egyszerűen csak tévézünk és pihenünk otthon*".

The English translation of this section (literal translation, retaining much of the Hungarian word order, grammar, and disfluencies):

In the mornings I usually get up about 7 o'clock. Then first I make my breakfast, this is the most important part, the highlight of the morning, so I spend the most time on it, therefore I need at least half an hour to have my breakfast, and only then come the other preparations like straighten up and so on... But I spend most of the day at school, after class I often go to the library, to complete my studies, um, well, um, yes. And, after school I often go for shopping, and then, um, after I get home, I cook for the next day, but it is not always so, but twice a week. Then if I have to prepare something for the next day, I do it and then we go out with my boyfriend, or go dance what I have already mentioned, or we go to the cinema or we just watch TV and have a rest at home.

These approximately one-minute long speech samples were taken as "original", and each was modified in four different ways. Using the program Audacity, we manually measured pauses that were longer than 100 ms (based on the Hungarian tradition (Gósy, 2003) Gósy and Kovács, 2008 and international studies (Henderson et al., 1966) (Schönpflug, 2008) and performed the following four manipulations to each: lengthening by 18%, lengthening by 50%, shortening by 21%, and shortening by 50%. Only one manipulation was performed on each sample. For

example, variations of speech samples with 21% shortened pauses were created so that all of the pauses that were longer than 100 ms in the original speech samples were shortened by 21%.

Modifications with 21% and 18% were based on our previous results (Szabó, 2008) where music and autobiographical recall were used to induce sad and happy emotional states. According to the results, the durations of silent pauses were 18% longer in a sad emotional state and 21% shorter in a happy emotional state compared to the neutral state. Based on observations in a pre-test, we also increased modifications by 50%.

Manipulations resulted in five variants for each of the five speech samples: variants with pauses elongated by 18% or 50%, variants with pauses shortened by 21% or 50%, and the original version. Thus, pauses became 50%, 79%, 100% (original), 118%, and 150% of the original versions (resulting in the following variants: 0.5, 0.7, 1 (original), 1.18, and 1.5.). (The description of stimuli is presented in Tables 1 and 2.).

To assess the naturalness of the materials, we conducted an additional rating study, where 75 university students (21 male and 54 female) with a mean age of 22.86 (SD = 4.50) differed from those who participated in the experiment, rated the naturalness of the sound files. In the same arrangement as in the experiment (see Table 3), five groups of participants listened to the speech samples and subjects were asked to rate how natural the sound files sounded on a scale of 1–6 (1: very artificial sounding, 6: absolutely natural sounding). Participants rated the naturalness of the sound files for an average of 4.66 (SD = 1.17), indicating that the material sounded quite natural and pause manipulation did not make the sounding artificial.

2.2. Participants

2.2.1. Hungarian sample

Fifty Hungarian raters (16 male and 34 female) participated in the first experiment, all of whom were students of the Budapest University of Technology and Economics. Participants were studying engineering or communications. They had a mean age of 20.12 years (SD = 1.41). All of these participants were native Hungarian speakers. Participants received university credits for their participation.

Table 1
Mean fundamental frequency, mean intensity, and the number of pauses of the original material.

Speaker	Mean fundamental frequency (SD) in Hz	Mean intensity (SD) in dB	Number of pauses
A	204.5 (8.3)	64.2 (0.7)	22
B	188.0 (20.4)	59.1 (2.1)	20
C	144.1 (19.7)	63.5 (1.7)	22
D	135.8 (11.3)	62.8 (1.4)	22
E	140.3 (9.4)	64.1 (1.3)	16

2.2.2. Austrian sample

Thirty-eight (8 male and 30 female) native German speaking Austrian subjects participated in the second experiment. Participants were all students of the Linguistic Institute at the University of Vienna, but twelve of them also studied other majors, such as history, pedagogy or languages. They had a mean age of 27.58 years (SD = 12.00). German was their native language, but three of them also spoke additional native languages. One subject spoke German and English, one spoke German and Dutch, and one spoke German, Polish, and Russian. With the exception of one participant, every subject spoke at least one foreign language. On average, participants spoke 2.74 foreign languages (SD = 1.14), but none of them understood or spoke Hungarian. Participation was voluntary, and subjects did not receive any compensation for their participation.

2.3. Procedure

Before the experiment began, participants were informed that the relationship between certain emotions and certain speech parameters would be studied. Participants were asked to judge five speech samples by filling out a questionnaire. After listening to each speech sample, participants were asked to rate how angry, sad, disgusted, happy, surprised, scared, positive, and heated the speaker seemed on a scale of 1–6 (not at all – very much). In addition to the basic emotions, *positive* and *heated* labels were believed to signal the valence and arousal dimensions. Austrian raters received questionnaires in German (the translation of the expressions was:*der Sprecher war: wütend, traurig, angeekelt, froh, überrascht, hatte Furcht, war positiv, war erregt*). Hungarian participants also rated the neutrality of the *content* of the speech samples on a scale of 1–6 (1: not at all emotionally-neutral content, 6: completely emotionally-neutral content). Thus, we tried to emphasise the fact that content and emotions can be independent of one another. In all cases, we emphasised that participants should rate the entire speech sample and should not decide too early in the utterance, and questionnaires must only be completed after listening to the full sample during the minute long pause between speech samples. Participants were instructed not to think too much about ratings, but to decide spontaneously.

One-minute long speech pauses were provided for answering the questions, and this time was intended to delete the "effect" of the speech that had been previously listened to so that participants had less opportunity to compare speech samples. Accordingly, listeners had to wait until the end of the one minute pauses. They were not allowed to fast-forward, even when they had answered all of the questions. All listeners participated in the experiment individually, and the speech samples were played on a computer via a headset. They answered the questions on a paper form questionnaire. The volume of the speech samples was the same for all participants, and the subjects were not allowed to adjust it. We previously tested the appropri-

Table 2
Length of the sound files, average length of pauses, total length of pauses, and pause ratio of the sound files across all 5 samples for each speaker.

Speaker	Length of sound file (s) Pause modification					Average length of pauses (ms) Pause modification					Total length of pauses (s) Pause modification					Pause ratio (%) Pause modification				
	*0.5	*0.79	*1	*1.18	*1.5	*0.5	*0.79	*1	*1.18	*1.5	*0.5	*0.79	*1	*1.18	*1.5	*0.5	*0.79	*1	*1.18	*1.5
A	52.4	55.7	58.0	60.0	63.5	253	402	506	597	759	5.6	8.8	11.1	13.1	16.7	10.6	15.9	19.2	21.9	26.3
B	45.0	48.6	51.1	53.3	57.2	306	486	612	722	918	6.1	9.7	12.2	14.4	18.4	13.6	20.0	24.0	27.1	32.1
C	59.1	62.4	64.4	66.6	70.1	250	396	499	589	749	5.5	8.7	11.0	13.0	16.5	9.3	14.0	17.0	19.5	23.5
D	50.7	54.0	56.3	58.2	61.6	250	397	500	590	750	5.5	8.7	11.0	13.0	16.5	10.8	16.2	19.6	22.3	26.8
E	44.3	47.2	49.1	50.9	53.8	302	480	605	714	907	4.8	7.7	9.7	11.5	14.5	10.9	16.3	19.7	22.4	26.9

ateness of the loudness of the sound files, based on the opinion of several persons. During the experiments all participants reported that they could hear and understand the utterances well.

2.4. Experimental arrangement

In both of the experiments, we had five experimental conditions. Each participant judged each speaker only once, and all of the speech samples he or she listened to were modified with a different pause ratio (experimental arrangement is seen in Table 3). Participants belonging to the same group listened to speech samples according to a "Balanced Latin Square" arrangement (Williams, 1949), which limits the risk of carryover effects With this method we aimed to avoid having one person listen to the same speech more than once.

3. Results of the first experiment: ratings of Hungarian participants

3.1. Differences between speakers

To evaluate whether speakers were rated as different regarding the emotionality of theirs speech's content, a univariate analysis of variance (ANOVA) using the GLM procedure was conducted with the original, non-manipulated speech samples of the speakers (5: *A*, *B*, *C*, *D*, *E*) as independent variable and the ratings on the *content neutrality* scale (1–6, where 1: not at all emotionally neutral, 6: completely emotionally neutral) as dependent variable. (We selected ratings on original speech samples from all of the answers.) The content of the original speech samples was rated by the Hungarian participants between 2.8 and 4.8, and ANOVA revealed a trend level result (main effect of speaker: $F(4,45) = 2.355$, $p = 0.068$, $\eta_p^2 = 0.173$), indicating that the five original speech samples were slightly different regarding the neutrality of their content.

In an analysis with the pause length manipulation (5 levels: 0.5, 0.79, 1, 1.18 vs. 1.5) as the independent variable and *content neutrality* ratings (1–6, where 1: not at all emotionally neutral, 6: completely emotionally neutral) as the dependent variable, we found that Hungarian participants rated the content of the speech samples between 3.64 and 4.24, and there was no significant difference between the five speech samples (main effect of pause length: ANOVA; $F(4,245) = 1.07$, $p = 0.37$, $\eta_p^2 = 0.017$). In conclusion, pause manipulation did not have an effect on content neutrality ratings.

We also analysed whether original samples were different regarding emotion ratings in an ANOVA with the non-manipulated speech samples of the speakers (5: *A*, *B*, *C*, *D*, *E*) as independent variable and ratings (1–6) on emotion scales (angry, sad, disgusted, happy, surprised, scared, positive, and heated) as the dependent variables. (Again, we selected ratings on original speech samples from all of the answers.) Separate ANOVAs for each emotion scale yielded significant differences on disgusted ($F(4,45) = 3.59$, $p = 0.013$, $\eta_p^2 = 0.242$), happy ($F(4,45) = 10.27$, $p < 0.001$, $\eta_p^2 = 0.477$), positive ($F(4,45) = 7.44$, $p < 0.001$, $\eta_p^2 = 0.398$), and heated ($F(4,45) = 5.28$, $p = 0.001$, $\eta_p^2 = 0.32$) ratings. Post-hoc LSD pairwise comparisons indicated that speaker *A* was rated as happier and more positive than other speakers and was less disgusted than speaker *C*. Speaker *E* was rated as the most heated and was significantly different from speakers *A*, *B*, *C*, and *D*. Again, speaker *E* was rated as more disgusted than speakers *A*, *B* and *D* (the exact*p* values are presented in Table 4). All in all, the original speech samples were not totally neutral from the perspective of content and emotional load, and listeners credited different emotional characteristics to them.

3.2. Relationship between pause modification and emotion judgments

To examine our research question, we investigated how pause conditions are related to judgments on the emotion

Table 3
Experimental arrangement. Letters (a–e) indicate the speech samples, while numbers indicate the rate of pause modification. Participants in group A listened to speech a with 50% abbreviated pauses (a0.5), speech b with 21% abbreviated pauses (b0.79), speech c in the original version (c1) etc.

Group A	Group B	Group C	Group D	Group E
a 0.5	a 1.5	a 1.18	a 1	a 0.79
b 0.79	b 0.5	b 1.5	b 1.18	b 1
c 1	c 0.79	c 0.5	c 1.5	c 1.18
d 1.18	d 1	d 0.79	d 0.5	d 1.5
e 1.5	e 1.18	e 1	e 0.79	e 0.5

scales. We performed univariate ANOVAs using the GLM procedure, and tested differences in mean ratings between the five pause modifications separately for each emotion scale (angry, sad, disgusted, happy, surprised, scared, positive, and heated). For all ANOVAs, the independent variable was pause length modification (5 levels: 0.5, 0.79, 1, 1.18 vs. 1.5), while the dependent variables were the ratings (1–6) given on the emotion scales. We used an alpha level of .05 for all ANOVAs and LSD post hoc tests.

The ANOVA yielded statistically significant differences on the sad scale and trend level differences on the happy scale (the results of the ANOVAs and mean ratings are presented in Table 5).

Post- hoc tests were used to investigate these differences in greater detail. LSD pairwise comparisons indicated that variants 1.5 were rated as significantly sadder than variants 0.5, 0.79, and 1; and variants 1.18 were rated as significantly sadder than variants 0.5. For the happy scale, there was a significant difference between variants 0.5 and variants 1.5, indicating that utterances with shorter pauses were rated as happier (the exact p values are presented in Table 6).

4. Results of the second experiment: ratings of Austrian participants

4.1. Differences between speakers

As with the Hungarian participants, we analysed whether Austrian raters judged Hungarian speakers' emotions differently. Univariate ANOVAs with original speech samples of the speakers (5: *A*, *B*, *C*, *D*, *E*) as independent variable and ratings (1–6) on each emotion scales (angry, sad, disgusted, happy, surprised, scared, positive, and heated) as dependent variable were conducted. (We selected ratings on original speech samples from all of the answers.) ANOVA yielded statistically significant differences for the sad ($F(4,33) = 5.04$, $p = 0.003$, $\eta_p^2 = 0.379$), disgusted ($F(4,32) = 5.23$, $p = 0.002$, $\eta_p^2 = 0.395$), happy ($F(4,32) = 14.73$, $p < 0.001$, $\eta_p^2 = 0.648$), scared ($F(4,32) = 2.86$, $p = 0.039$, $\eta_p^2 = 0.263$), and positive ($F(4,32) = 15.32$, $p < 0.001$, $\eta_p^2 = 0.657$) scales.

Post-hoc LSD pairwise comparisons indicated that the Austrians considered speakers *A* and*E* as happier and more positive than speakers *B, C*, and *D*, and speaker *A* was rated as even more positive and happier than speaker *E*. In line with this finding, *A* and *E* were rated as less sad than speakers *B, C*, and *D*. Speaker *B* was rated as more disgusted than other speakers and more scared than speakers *A* and *E* (the exact p values are presented in Table 7). In sum, Austrian participants ascribed different emotional characteristics to the original speech samples.

4.2. Relationship between pause modification and emotion judgments

A univariate ANOVAs with pause length modification (5 levels: 0.5, 0.79, 1, 1.18 vs. 1.5) as the independent variable and the Austrians' ratings (1–6) on the emotion scales (angry, sad, disgusted, happy, surprised, scared, positive, and heated) as the dependent variables yielded no statistically significant effects. Trend level differences for the happy and positive scales were observed (the results of the ANOVAs and mean ratings are presented in Table 8).

To explore the differences in more detail, post hoc tests were made. LSD pairwise comparisons indicated that variants 1.5 were rated as significantly less happy than variants 0.5, 1, and 1.18, and variants 1.5 were rated as significantly less positive than all the other variants (the exact p values are presented in Table 9). The results indicate that utterances with shorter pauses were rated as happier and more positive.

5. Comparing the ratings of the two groups of participants

To test whether there is a difference between the two groups regarding emotion judgments or there is an interaction between the independent variables, we performed a MANOVA with language (2 levels: Hungarian, German), pause modification (5 levels: 0.5, 0.79, 1, 1.18 vs. 1.5), and speaker (5 levels: *A*, *B*, *C*, *D*, *E*) as independent variables, and ratings (1–6) on emotion scales (angry, sad, disgusted, happy, surprised, scared, positive, and heated) as dependent variables. MANOVA revealed significant main effect of language ($F(8,378) = 14.13$, $p < 0.001$, $\eta_p^2 = 0.230$), pause modification $F(32,1524) = 1.47$, $p < 0.05$, $\eta_p^2 = 0.030$), and speaker ($F(32,1524) = 8.38$, $p < 0.001$, $\eta_p^2 = 0.150$). In addition, there was an interaction of language and speaker ($F(32,1524) = 3.01$, $p < 0.001$, $\eta_p^2 = 0.059$). No other interactions were significant.

The effect of language was significant for the angry ($F(1,385) = 17.11$, $p < 0.001$, $\eta_p^2 = 0.043$), the sad ($F(1,385) = 15.61$, $p < 0.001$, $\eta_p^2 = 0.039$), the surprised ($F(1,385) = 17.13$, $p < 0.001$, $\eta_p^2 = 0.043$), the scared ($F(1,385) = 22.28$, $p < 0.001$, $\eta_p^2 = 0.055$), the positive ($F(1,385) = 4.07$, $p < 0.05$, $\eta_p^2 = 0.010$) and the heated ($F(1,385) = 73.45$, $p < 0.001$, $\eta_p^2 = 0.16$) scales. Austrian participants rated Hungarian speakers as angrier, sadder, more surprised, more scared, less positive, and more heated than did Hungarian raters (see Fig. 1).

The main effect of pause modification was significant for the sad ($F(4,385) = 4.71$, $p < 0.05$, $\eta_p^2 = 0.047$), the happy ($F(4,385) = 4.03$, $p < 0.05$, $\eta_p^2 = 0.040$), the scared ($F(4,385)$=4.28, $p < 0.05$, $\eta_p^2 = 0.043$), and the positive ($F(4,385) = 3.16$, $p < 0.05$, $\eta_p^2 = 0.032$) scales. Post hoc LSD pairwise comparisons indicated that variants 1.5 were rated as significantly sadder than variants 0.5, 0.79, and 1; and variants 1.18 were rated as significantly sadder than variants 0.5, 0.79. Variants 1.5 were rated as significantly less happy and less positive than the other variants. Variants 1.5 were rated as more scared than the other variants of pause modification. In sum, utterances with longer pauses were rated as sadder, less happy, more scared and less positive (the exact p values are presented in Table 10).

Table 4
Differences between speakers: significant effects on certain emotion scales, the direction of the relation, and exact p values from significant LSD multiple comparisons (Hungarians' ratings).

Emotion scale	Speaker	Direction of the relationship	Speaker	p value
Disgusted	*C*	>	*A*	0.026
	E	>	*A*	0.003
		>	*B*	0.017
		>	*D*	0.011
Happy	*A*	>	*B*	<0.001
		>	*C*	<0.001
		>	*D*	<0.001
		>	*E*	<0.001
Positive	*A*	>	*B*	<0.001
		>	*C*	0.001
		>	*D*	0.001
		>	*E*	<0.001
Heated	*E*	>	*A*	0.001
		>	*B*	0.008
		>	*C*	0.002
		>	*D*	<0.001

Given that there was no interaction between language and pause modification, the results indicate that pause length plays a similar role for the Hungarian and the German speaking raters in ascribing sad, happy, scared and positive emotional states, and longer pauses indicate sadder, less happy, more scared and less positive emotional states. (Mean ratings of the Hungarians and the Austrians are seen in Fig. 2.).

The effect of speaker was analysed more detailed separately for the two language groups (see Sections 3.1 and 4.1). The speaker and language interaction was significant on the angry ($F(4,385) = 4.21$, $p < 0.05$, $\eta_p^2 = 0.042$), sad ($F(4,385) = 5.32$, $p < 0.001$, $\eta_p^2 = 0.052$), disgusted ($F(4,385) = 5.6$, $p < 0.001$, $\eta_p^2 = 0.055$), scared ($F(4,385) = 5.67$, $p < 0.001$, $\eta_p^2 = 0.056$), positive ($F(4,385) = 5.33$, $p < 0.001$, $\eta_p^2 = 0.052$), and heated scales ($F(4,385) = 4.61$, $p < 0.05$, $\eta_p^2 = 0.046$). Post hoc LSD pairwise comparisons indicated that Austrians rated the speakers as angrier, sadder, more surprised, more scared, less positive, and more heated, in the case of speaker *E* and emotions angry, sad, and positive, ratings were in the opposite direction. Moreover, speaker *A* and *E* were rated as similarly scared, and *E* was rated as similarly heated by the two groups ($ps < 0.05$).

6. Discussion

The aim of the study was to investigate how changes in only one parameter influence the emotion judgments of listeners. Trends in the data showed that both in the Hungarian and Austrian groups, the length of silent pauses influenced listeners in attributing emotional states to the speaker. In the case of the Hungarians, the effect of pause length was most evident on the sad scale, and pause length manipulation had a significant effect on emotion ratings. The same speech samples were rated as sadder when pauses were longer. There was a smaller, trend level effect on the happy scale, indicating that listeners rate slightly differently the same speech samples with different pause length in generally, variants that contained shorter pauses were rated as significantly happier that variants that contained longer pauses. In the case of the German speaking raters, the pause length manipulation had a general trend level effect on the happy and positive scales. Those variants which pauses were elongated by 50% were rated as significantly less happy than those variants where the pauses were shortened by 50% or were elongated by 18% or the original versions. The variants, containing the 50% elongated pauses were rated as significantly less positive than all the other variants.

Combining the results from the Hungarian and the Austrian raters, we found that pause length modification has a significant main effect on the sad, happy, scared, and positive scales. Given that there was no interaction between language and pause modification, the results indicate that pause length plays a similar role for the Hungarian and the German speaking raters in ascribing sad, happy, scared and positive emotional states; longer pauses indicate sadder, less happy, more scared and less positive emotional state. The strongest effect was observed in the case of the sad emotion by the Hungarians. The results are consistent

Table 5
Mean ratings and standard deviations of Hungarian participants on emotion scales, regarding pause manipulation and ANOVA summary.

Emotion scale	Mean ratings on emotion scales (SD) Pause manipulation					Df effect	Df error	F	p	η_p^2
	0.5	0.79	1	1.18	1.5					
Angry	1.46 (0.9)	1.46 (0.9)	1.68 (1.2)	1.62 (1.0)	1.50 (0.9)	4	245	0.49	0.743	0.008
Sad	2.00 (1.2)	2.2 (1.3)	2.32 (1.4)	2.58 (1.6)	2.96 (1.6)	4	244	3.33	0.011*	0.052
Disgusted	1.62 (1.1)	1.82 (1.3)	2.16 (1.4)	2.14 (1.2)	2.12 (1.3)	4	245	1.79	0.131	0.028
Happy	2.86 (1.4)	2.6 (1.6)	2.46 (1.4)	2.3 (1.3)	2.1 (1.3)	4	245	2.01	0.094	0.032
Surprised	1.66 (1.1)	1.70 (1.1)	1.44 (1.0)	1.78 (1.3)	1.90 (1.3)	4	245	1.07	0.373	0.017
Scared	1.52 (0.9)	1.58 (1.2)	1.56 (1.0)	1.82 (1.3)	2.08 (1.6)	4	245	1.87	0.117	0.030
Positive	3.28 (1.5)	3.14 (1.6)	3.06 (1.5)	2.84 (1.5)	2.62 (1.4)	4	245	1.52	0.198	0.024
Heated	2.24 (1.3)	2.14 (1.3)	2.10 (1.3)	2.02 (1.2)	1.68 (0.9)	4	245	1.59	0.178	0.025

* ANOVA is significant at the 0.05 level.

Table 6
Differences in sad and happy emotion ratings regarding pause manipulation: the direction of the relation and exact p values from significant LSD multiple comparisons (Hungarians' ratings).

	Variant	Direction of the relationship	Variant	p value
Sad	1.18	>	0.5	0.044
	1.5	>	0.5	0.001
		>	0.79	0.009
		>	1	0.027
Happy	0.5	>	1.5	0.008

with Breitenstein et al. (2001), Fairbanks and Hoaglin (1941) and Szabó (2008) studies which found that speech rate and speech pauses are strongly related to happy and sad emotional states.

With respect to the fear ratings, we found that with the increase of pause length, speakers were rated as being more scared. As fear and anxiety are interconnected emotions (Fontaine et al., 2007), our result can be related to previous studies on anxiety (Eldred and Price, 1958; Kasl and Mahl, 1965; Mahl, 1956; Pope et al., 1970; Hofmann et al., 1997; Laukka et al., 2008), which found that in speech during anxious situations or in the speech of social phobics, the number of pauses increased. Our results seem to contradict those of Breitenstein et al. (2001), who found that with the increase in speech rate, speakers were rated as more scared. However, in Bretensteins' experiment, the tempo was modified by slowing down and speeding up sentences in a supposedly continuous way while also modifying the articulation rate and speech rate. In our case, tempo modification was the result of just manipulating pause length and leaving articulation fixed. Thus, the results of the two studies do not necessarily contradict one another, but rather they are compatible: fear/anxiety causes a faster articulation rate (Fairbanks and Hoaglin, 1941) *and,* at the same time, longer pauses in speech. When these speech features are perceived, fear is ascribed to the speaker. Therefore, according to the results speech rate and the speech/pause ratio seem to related to happy-sad (valence dimension, see (Fontaine et al., 2007) and scared emotional states. When we are happy, sad, or scared, we show our emotional state to listeners by modifying the pause length in our speech. Furthermore, when with a person we are speaking to perceives changes in the speech/pause ratio, he or she will ascribe these emotional states to us. In the case of the other emotions, there was no observable effect of the pause manipulation. Probably, speakers tend to express anger, surprise, disgust, and heatedness by using shorter utterances and pauses play no role.

Table 7
Differences between speakers: significant effects on certain emotion scales, the direction of the relation, and exact p values from significant LSD multiple comparisons (Austrians ratings).

Emotion scale	Speaker	Direction of the relationship	Speaker	p value
Sad	*A*	<	*B*	0.007
		<	*C*	0.003
		<	*D*	0.017
	E	<	*B*	0.007
		<	*C*	0.003
		<	*D*	0.017
Disgusted	*B*	>	*A*	<0.001
		>	*C*	0.022
		>	*D*	0.001
		>	*E*	0.002
Happy	*A*	>	*B*	<0.001
		>	*C*	<0.001
		>	*D*	<0.001
		>	*E*	0.016
	E	>	*B*	0.002
		>	*C*	0.001
		>	*D*	0.001
Positive	*A*	>	*B*	<0.001
		>	*C*	<0.001
		>	*D*	<0.001
		>	*E*	0.008
	E	>	*B*	0.001
		>	*C*	0.001
		>	*D*	0.007
Scared	*B*	>	*A*	0.010
		>	*E*	0.007

Speech samples were produced by Hungarian volunteers and not by actors, as the role of the pauses in natural speech was the focus of the study. Though the original speech samples were not totally neutral regarding their content or emotional load, the experimental arrangement (Latin square and balanced Latin square, eliminating order effect) allowed us to use pause length modification as the only independent variable. Thus, the results of the statistical analysis indicated that differences in emotion judgments were only affected by pause length modification. However, we have to stress it here that when interpreting the results, it has to be taken into account that effect sizes were small and the LSD post hoc tests used in this experiment are considered as quite liberal.

The fact that there was no big difference between the emotion ratings that were given on certain variants is not surprising, as the topic of all speech samples was neutral, and intonation, loudness, articulation rate and other speech parameters corresponded with this emotionally neutral content. In addition, the rate of pause modification was not high. When listening to the variants of the same sound file one after another, differences are hard to detect. Furthermore, in the additional rating study, participants rated the sound files for quite natural sounding and none of the participants mentioned that the speech samples were artificial or strange.

In the experimental instructions, we emphasised to both groups that the task was to judge the entire speech sample because we assumed that modifications in speech pauses only have an effect during longer passages. Manipulating the length of speech pauses is not only interesting in itself, but also determines the speech-pause ratio. Thus, not only the length of pauses followed by one another, but also the ratio of pauses to the whole speech is what listeners perceive and what influences emotion ascribing.

Table 8
Mean ratings and standard deviations of Austrian participants on emotion scales, regarding pause manipulation, and ANOVA summary.

Emotion scale	Mean ratings on emotion scales (SD) Pause manipulation					Df effect	Df error	F	p	η_p^2
	0.5	0.79	1	1.18	1.5					
Angry	1.95 (1.2)	2.00 (0.9)	1.89 (1.2)	1.76 (0.9)	2.08 (1.2)	4	184	0.448	0.773	0.010
Sad	2.74 (1.7)	2.63 (1.6)	2.71 (1.7)	3.18 (1.8)	3.47 (1.8)	4	185	1.738	0.143	0.036
Disgusted	1.82 (1.3)	2.16 (1.4)	1.97 (1.3)	1.74 (1.1)	1.87 (0.9)	4	184	0.713	0.584	0.015
Happy	2.53 (1.5)	2.37 (1.3)	2.65 (1.7)	2.47 (1.6)	1.74 (1.4)	4	184	2.14	0.077	0.045
Surprised	2.13 (1.4)	2.00 (1.1)	2.27 (1.4)	2.29 (1.4)	2.18 (1.3)	4	184	0.291	0.883	0.006
Scared	2.11 (1.4)	2.00 (1.3)	2.11 (1.4)	2.32 (1.4)	2.79 (1.7)	4	184	1.859	0.120	0.039
Positive	2.87 (1.6)	2.82 (1.6)	2.97 (1.7)	2.97 (1.7)	2.05 (1.6)	4	184	2.16	0.076	0.045
Heated	3.21 (1.4)	2.97 (1.3)	3.14 (1.5)	2.63 (1.2)	3.11 (1.4)	4	183	1.052	0.382	0.022

Table 9
Differences in happy and positive emotion ratings regarding pause manipulation: the direction of the relation and exact p values from significant LSD multiple comparisons (Austrians' ratings).

	Variant	Direction of the relationship	Variant	p value
Happy	1.5	<	0.5	0.023
		<	1	0.009
		<	1.18	0.034
Positive	1.5	<	0.5	0.030
		<	0.79	0.042
		<	1	0.015
		<	1.18	0.015

Table 10
Differences in emotion ratings regarding pause manipulation: the direction of the relation and exact p values from significant LSD multiple comparisons (results from the Hungarian and the Austrian raters together).

Emotion scale	Variant	Direction of the relationship	Variant	p value
Sad	1.18	>	0.5	0.019
		>	0.79	0.036
	1.5	>	0.5	<0.001
		>	0.79	<0.001
		>	1	0.001
Happy	1.5	<	0.5	<0.001
		<	0.79	0.004
		<	1	0.002
		<	1.18	0.020
Scared	1.5	>	0.5	0.001
		>	0.79	0.001
		>	1	0.001
		>	1.18	0.042
Positive	1.5	<	0.5	0.001
		<	0.79	0.004
		<	1	0.002
		<	1.18	0.018

Hungarian and Austrian subjects participated in our research. Previous intercultural/interlinguistic emotion recognition studies (Vanbezooijen et al., 1983; Thompson and Balkwill, 2006; Scherer et al., 2001; Albas et al., 1976; Pell and Skorup, 2008; Fónagy and Magdics, 1963) have found that people can identify emotions from foreign languages with better than chance accuracy. At the same time, individuals are more accurate with regard to sentences uttered in their native language than those in a foreign language (Beier and Zautra, 1972; Vanbezooijen et al., 1983; Scherer et al., 2001). Based on our results, emotion ascribing works in a similar manner for Hungarians and Austrians with a German native language when the length and rate of speech pauses change. In our opinion, the relationship between speech pauses and emotions might be universal, but more studies are needed to confirm this hypothesis. A good theory to explain the similar results is the Elfenbein–Ambady cultural proximity hypothesis (Elfenbein and Ambady, 2003), which argues that people from similar cultures are similar in coding emotions, and thus decode one another's emotions more easily than people from more distant cultures. Hungarian and Austrian cultures were often interconnected throughout history, and because they are neighbouring countries, they are also geographically close to one another. Thus, although the German and Hungarian languages are very different (they belong not even into the same language family), the function of speech pauses in emotion expression might be the same, and longer pauses shift answers to the *sadder* and *less happy, less positive* and*more scared* direction.

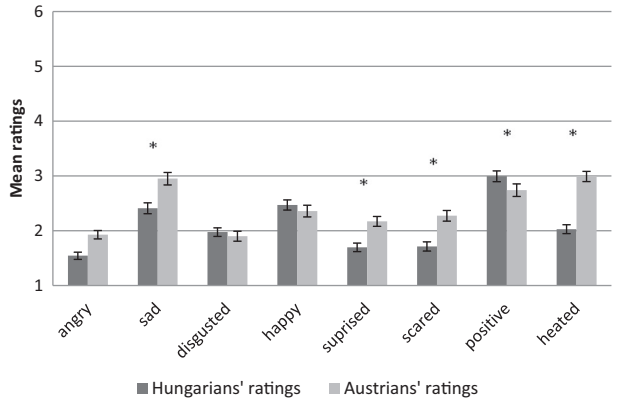

Fig. 1. Hungarians' and Austrians' mean ratings on the emotion scales across all 5 samples for each speaker (significant differences are marked by *) Error bars represent the standard error.

At the same time, we found differences between the two groups. First, the two groups differed in the emotion judgments of speech samples. Based on the statistical analysis, Austrian participants rated the Hungarian speakers as

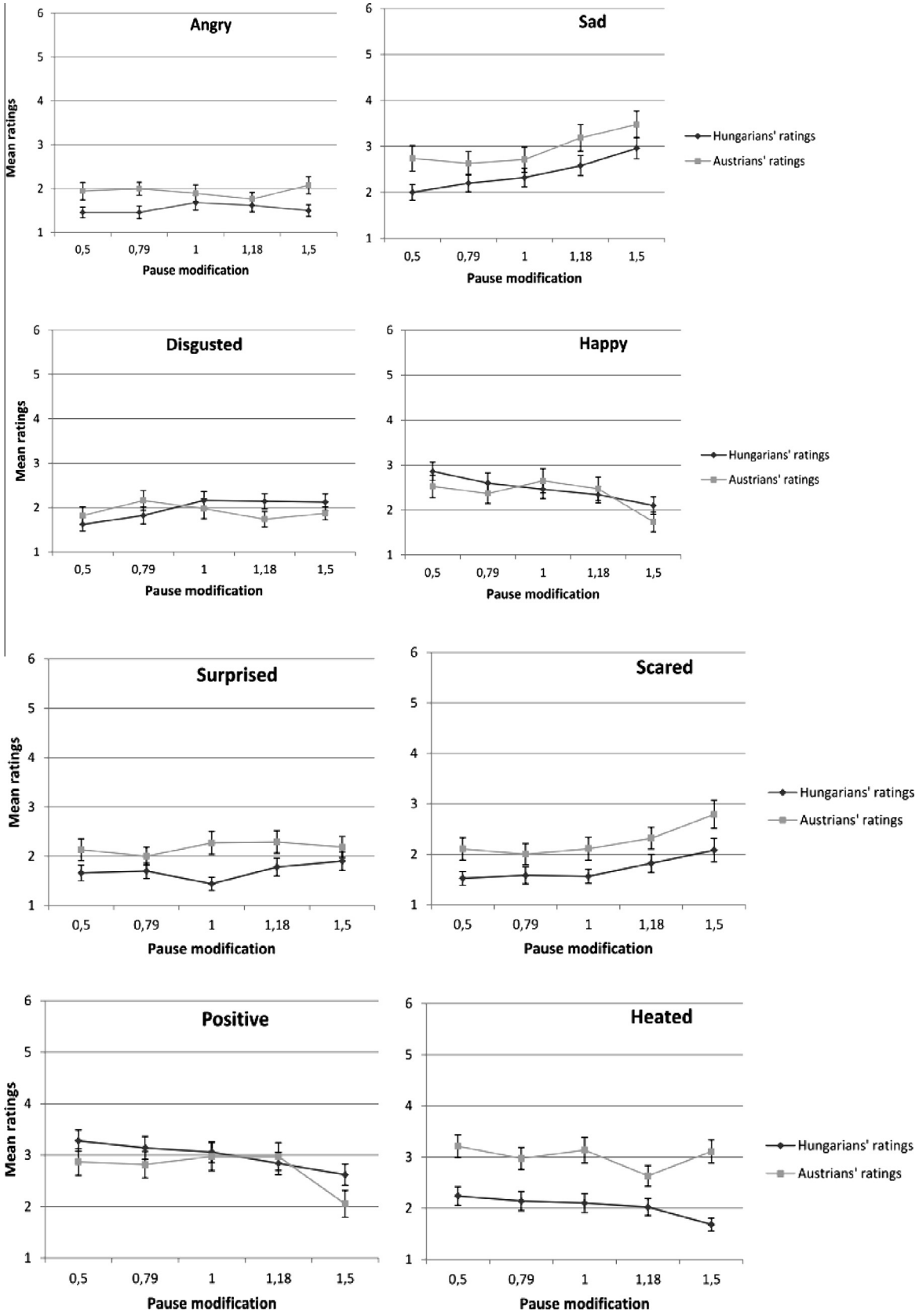

Fig. 2. Mean ratings of Hungarian and Austrian participants on the emotion scales, divided by pause modification. Error bars represent the standard error.

angrier, sadder, more surprised, more scared, less positive, and more heated than did Hungarian raters. The results cannot be connected with the valence or the arousal dimensions, as although speakers generally were rated by the Austrians as less positive, but one exception is the surprised scale, and more heated yet again, an exception is the sad scale. (see Fontaine et al., 2007). It is conceivable that the results were affected by language differences. Thus, speech parameters such as loudness, pitch, and melody and the common appearance of these factors made the Austrians raters, compared to the Hungarian ones, feel more like that the Hungarian speakers were angrier, sadder, more surprised, scared, negative, and heated. Of course, results might also have been influenced by the fact that the sentences were meaningful for the Hungarian raters, while Austrians could only base their rating on the vocal and temporal features.

Speakers' voices were judged differently by Hungarian and Austrian participants as well, and for several speakers, higher ratings were given on certain emotion scales. Judgments of the two groups were only partly similar, as speaker *A* was rated as happier and as more positive than the other speakers. As speaker *A* had the highest pitch and high pitch reveals positive emotional states no matter of culture, is likely that this effect is responsible for this similarity. In sum, individual characteristics of speech and the linguistically and culturally determined way of emotion expression fundamentally influence the attributes and emotional states we ascribe to speakers.

Our work highlights the role of speech pauses in emotion expression and recognition and stresses the importance of the topic. Our results show that speech pauses play an important role in emotion ascribing, which might be a regional/cultural, if not universal, phenomenon. To confirm this assertion, an "inversion" of the experiment would be needed, or in other words, manipulating pauses in German speech samples and getting them rated by subjects with a German and Hungarian mother tongue.

Acknowledgements

We thank Florian Menz, professor at the Linguistic Institute, the University of Vienna for helping in providing conditions for the experiment. This work was supported by a scholarship from the Austria-Hungary Action Foundation to the first author. We thank Csaba Szabó, Ágnes Lukács, Dezső Németh, Karolina Janacsek, and István Winkler for valuable feedback and suggestions.

References

Albas, D.C., Mccluskey, K.W., Albas, C.A., 1976. Perception of emotional content of speech – comparison of 2 Canadian groups. Journal of Cross-Cultural Psychology 7 (4), 481–490.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70 (3), 614–636.

Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech Communication 46 (3–4), 252–267.

Beaupre, M.G., Hess, U., 2005. Cross-cultural emotion recognition among Canadian ethnic groups. Journal of Cross-Cultural Psychology 36 (3), 355–370.

Beier, E.G., Zautra, A.J., 1972. Identification of vocal communication of emotions across cultures. Journal of Consulting and Clinical Psychology 39 (1), 166.

Bergmann, G., Goldbeck, T., Scherer, K.R., 1988. Emotional impression of prosodic speech markers. Zeitschrift für experimentelle und angewandte Psychologie 35 (2), 167–200.

Biehl, M. et al., 1997. Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): reliability data and cross-national differences. Journal of Nonverbal Behavior 21 (1), 3–21.

Breitenstein, C. et al., 1996. Erfassung der Emotionswahrnehmung bei zentralnervösen Läsionen und Erkrankungen: Psychometrische Gütekriterien der "Tübinger Affekt Batterie". Neurologie and Rehabilitation 2 (93–101).

Breitenstein, C., Van Lancker, D., Daum, I., 2001. The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. Cognition and Emotion 15 (1), 57–79.

Burkhardt, F., Sendlmeier, W.F., 2000. Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis. In: ISCA Workshop (ITRW) on Speech and Emotion 2000, Belfast.

Burkhardt, F. et al., 2006. Emotional prosody – does culture make a difference? In: Proceedings of Speech Prosody 2006, Dresden, Germany.

Cahn, J.E., 1990. The generation of affect in synthesized speech. Journal of the American Voice I/O Society 8, 1–19.

Carlson, R., 1992. Synthesis – modeling variability and constraints. Speech Communication 11 (2–3), 159–166.

Deppermann, A., Lucius-Hoene, G., 2005. *Trauma erzählen – kommunikative, sprachliche und stimmliche Verfahren der Darstellung traumatischer Erlebnisse*. Psychotherapie und Spzialwissenschaft. Zeitschrift für Qualitative Forschung und klinische Praxis 1, 35–73.

Ekman, P., Friesen, W.V., 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology 17 (2), 124–129.

Ekman, A., Strom, J., 1969. Habit forming drugs (7): somatic complications. Lakartidningen 66 (48), 5021–5024.

Ekman, P., Sorenson, E.R., Friesen, W.V., 1969. Pan-cultural elements in facial displays of emotion. Science 164 (3875), 86–88.

Ekman, P. et al., 1987. Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology 53 (4), 712–717.

Eldred, S.H., Price, D.B., 1958. A linguistic evaluation of feeling states in psychotherapy. Psychiatry 21 (2), 115–121.

Elfenbein, H.A., Ambady, N., 2003. Cultural similarity's consequences – a distance perspective on cross-cultural differences in emotion recognition. Journal of Cross-Cultural Psychology 34 (1), 92–110.

Elfenbein, H.A. et al., 2007. Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. Emotion 7 (1), 131–146.

Fairbanks, G., Hoaglin, L.W., 1941. An experimental study of the durational characteristics of the voice during the expression of emotion. Speech Monographs 8 (1), 85–90.

Fónagy, I., Magdics, K., 1963. Emotional patterns in intonation and music. Zeitschrift fur Phonetik 16, 293–326.

Fontaine, J.R. et al., 2007. The world of emotions is not two-dimensional. Psychological Science 18 (12), 1050–1057.

Goldman-Eisler, F., 1958. Speech production and the predictability of words in context. Quarterly Journal of Experimental Psychology 10, 96–106.

Goldman-Eisler, F., 1968. Psycholinguistics: Experiements in Spontaneous Speech. Academic Press, New York.

Gósy, M., 2003. A spontán beszédben előforduló megakadásjelenségek gyakorisága és összefüggései. Magyar Nyelvőr 127, 257–277.

Gósy, M., 2008. Magyar spontán beszéd adatbázis – BEA. Beszédkutatás 2008, 116–128.
Gósy, M., Kovács, M., 2008. Virtual sentences of spontaneous speech. In: Human Factors and Voice Interactive Systems 2008, Springer, New York, London, pp. 193–207.
Henderson, A.I., Goldman-Eisler, F., Skarbek, A., 1966. Sequential temporal patterns in spontaneous speech. Language and Speech 9, 207–216.
Hofmann, S.G. et al., 1997. Speech disturbances and gaze behavior during public speaking in subtypes of social phobia. Journal of Anxiety Disorders 11 (6), 573–585.
Izard, C.E., 1971. The Face of Emotion. Appleton-Century-Crofts, New York.
Izard, C.E., 1994. Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychological Bulletin 115 (2), 288–299.
Johnson, W.F. et al., 1986. Recognition of emotion from vocal cues. Archives of General Psychiatry 43 (3), 280–283.
Jovicic, S.T. et al., Serbian Emotional Speech Database: Design, Processing and Evaluation. In: SPECOM-20042004, p. 77–81.
Juslin, P.N., Laukka, P., 2001. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion 1 (4), 381–412.
Juslin, P.N., Scherer, K.R., 2005. Vocal expression of affect. In: Harrigan, J.A., Rosenthal, R., Scherer, K.R. (Eds.), The New Handbook of Methods in Nonverbal Behavior Research. Oxford University Press, New York, pp. 65–135.
Kasl, S.V., Mahl, G.F., 1965. The relationship of disturbances and hesitations in spontaneous speech to anxiety. Journal of Personality and Social Psychology 95, 425–433.
Ladd, D.R. et al., 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. Journal of the Acoustic Society of America 78 (2), 435–444.
Laukka, P. et al., 2008. In a nervous voice: acoustic analysis and perception of anxiety in social phobics' speech. Journal of Nonverbal Behavior 32 (4), 195–214.
Mahl, G.F., 1956. Disturbances and silences in the patient's speech in psychotherapy. Journal of Abnormal Psychology 53 (1), 1–15.
Matsumoto, D., 1993. Ethnic-differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. Motivation and Emotion 17 (2), 107–123.
Mccluskey, K.W., Albas, D.C., 1981. Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. International Journal of Psychology 16 (2), 119–132.
Mccluskey, K.W. et al., 1975. Cross-cultural differences in perception of emotional content of speech – study of development of sensitivity in canadian and mexican children. Developmental Psychology 11 (5), 551–555.
Pell, M.D., Skorup, V., 2008. Implicit processing of emotional prosody in a foreign versus native language. Speech Communication 50 (6), 519–530.
Pell, M.D. et al., 2009. Factors in the recognition of vocally expressed emotions: a comparison of four languages. Journal of Phonetics 37 (4), 417–435.
Pope, B. et al., 1970. Anxiety and depression in speech. Journal of Consulting and Clinical Psychology 35 (1), 128–133.
Rochester, S.R., 1973. The significance of pauses in spontaneous speech. Journal of Psycholinguistic Research 2 (1), 51–81.
Scherer, K.R., 2003. Vocal communication of emotion: a review of research paradigms. Speech Communication 40 (1–2), 227–256.
Scherer, K.R., Ladd, D.R., Silverman, K.E.A., 1984. Vocal cues to speaker affect – testing 2 models. Journal of the Acoustical Society of America 76 (5), 1346–1356.
Scherer, K.R., Banse, R., Wallbott, H.G., 2001. Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-Cultural Psychology 32 (1), 76–92.
Schönpflug, U., 2008. Pauses in elementary school children's verbatim and gist free recall of a story. Cognitive Development 23, 385–394.
Schröder, M., 2003. Experimental study of affect bursts. Speech Communication 40 (1–2), 99–116.
Szabó, E., 2008. A szomorú és a vidám érzelmi állapot megjelenése a beszédben. Magyar Pszichológiai Szemle 63 (4), 651–668.
Thompson, W.F., Balkwill, L.L., 2006. Decoding speech prosody in five languages. Semiotica 158 (1–4), 407–424.
Vanbezooijen, R., Otto, S.A., Heenan, T.A., 1983. Recognition of vocal expressions of emotion – a 3-nation study to identify universal characteristics. Journal of Cross-Cultural Psychology 14 (4), 387–406.
Williams, E.J., 1949. Experimental designs balanced for the estimation of residual effects of treatments. Australian Journal of Scientific Research Series A 2, 149–168.