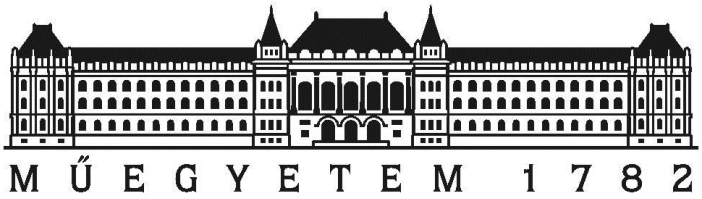

Budapest University of Technology and Economics, Faculty of Natural Sciences

PhD School in Psychology

Investigation of speech features of emotional states and of altered states of consciousness

PhD thesis booklet

Eszter Tisljár-Szabó

Supervisor: Dr. Csaba Pléh

2014

Introduction

Several models and experimental evidences exist which describe speech production or speech perception (e. g. Garrett, 1988; Levelt, 1989; Marslen-Wilson, Tyler, 1980). However, it is less known about the changes of speech in different than normal or neutral states (such as in emotional states or in altered states of consciousness).

We have everyday experiences about the way of speaking if someone is happy as something good has happened to him or he is anxious during a hard oral exam. We also know that we reasonably well can recognize the emotional or the alcohol influenced state of our friends only by listening theirs voice Still, it is an important research question what does exactly change in our speech in different kind of psychological states and what does influence the listener when he attributes any emotional or mental state to the speaker.

As a theoretical background, we have used Scherer's model of vocal emotion expression and perception (see Figure 1) (Scherer, 2003, 2013). According to the model, the process begins with the encoding of emotional speaker states when the emotional arousal of the speaker is accompanied by changes in speech in such a way as to produce emotion-specific patterns of acoustic parameters. These acoustic changes are called distal cues (distal in the sense of distant from the observer) and are transmitted by the speech signal to the listener. These perceived cues are the so called proximal cues (close to the observer).

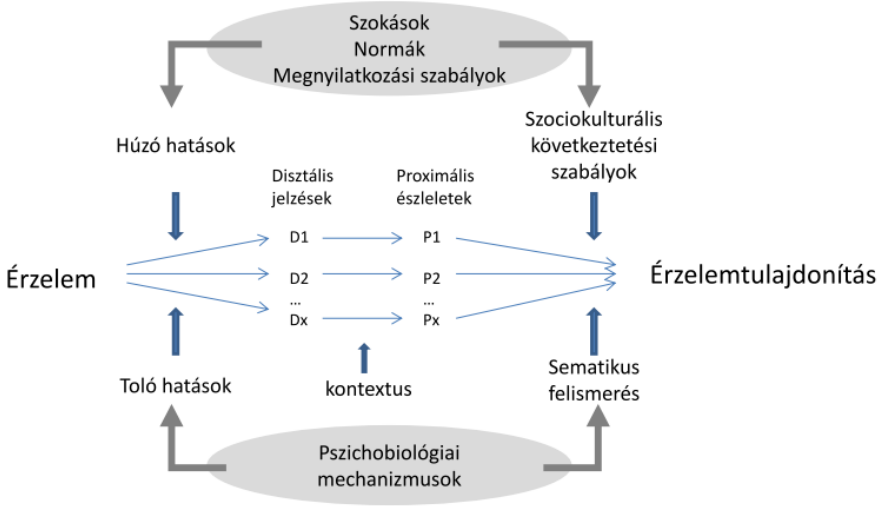

Figure 1. Scherer's model of vocal emotion expression and perception (based on Scherer, 2003, 2013; Grandjean, 2006)

The model highlights the fact that objectively measured distal characteristics are not necessarily equivalent to the proximal cues they produce in the observer. While the proximal cues are based on distal characteristics, the latter might be modified by the transmission channel (e.g. distance, noise) or the structural characteristics of the perceptual organ (e.g. selective enhancement of certain frequency bands). The result of the decoding process is an inference of speaker emotions and personality based on changes in perceived speech, the proximal cues.

As mentioned, an error can occur everywhere in the process. For example, it is possible that the respective emotional state does not produce reliable externalizations in the form of specific distal cues in the voice (e.g. because of hiding the real emotional state). It also can happen that although distal cues express speaker's emotional state, but are modified during transmission and perception in such a way that they no longer carry the essential information when they are proximally perceived by the listener (e.g. because of the noise). In addition, it is possible that also distal and proximal cues overlap, the inference mechanism "goes wrong", for example when ascribing different emotional state to the speaker as it was intended to express, while listening to foreign language speech.

As Figure 1 shows, according to the model, both emotion expression (coding) and perception (decoding) are influenced by psychobiological mechanisms (push effects) and norms, sociocultural rules (pull effects).

While the mentioned phenomena usually occur together in social situations, they need to be separated when they are studied. If researcher is interested in the question of the co-occurrences of certain inner states and speech features he needs to plan a 'production study' while if he wants to know more about the configuration of speech parameters that influence listeners in ascribing emotional states to the speaker, a 'perception study' is needed. One also can investigate the role of only one speech parameter in emotion ascribing. While doing this, one systematically manipulates distal cues to clearly establish the causal structure of the inference. (About the methods in more detail see Sudy [1].)

Connected to the topic of the dissertation, we planned researches to investigate the proximal and the distal cues. In the first case, we investigated in native and in foreign language speech how speech pause length influences how listeners ascribe emotional states to the speaker. Only a few studies have investigated the relation of speech pauses and emotional states. Previous production research showed that number and length of silent pauses increase when the speaker is sad or anxious (Deppermann és Lucius-Hoene, 2005; Szabó, 2008; Mahl, 1956;

Pope et al., 1970; Hofmann et al., 1997; Laukka et al. 2008). Further, a perception study (Breitenstein et al., 2001) investigated how manipulation of speech rate changes the emotion ratings attributed to the speaker by the listener. Regarding the results, in most of the sentences, if the utterance became slower, then the frequency of sad ratings increased.

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) showed that listeners can accurately detect and categorize emotional states from speakers with a different language and cultural background. 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.

Thus, in our 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 neutral speech samples. Secondly, we investigated whether there is a pause length effect when listeners with a German mother tongue listen to Hungarian speech.

To investigate the distal cues, in our second research we investigated the effect of acute alcohol consumption on speech and on short term memory function. The previous studies about the topic are contradicting. Some studies showed that in an alcohol influenced state short term memory performance decreases (Miller és Dolan, 1974; Maylor és Rabbitt, 1987; Nordby et al., 1999), others showed no effect of alcohol (Jones és Jones, 1980), while others again (Baum-Baicker, 1985) even found an increased performance in an intoxicated state. Regarding the effect of alcohol on speech, results are inconsistent as well. Most of the studies showed that in an intoxicated state the length of pauses increases (Künzel et al. 1992), the speech rate decreases (Behne et al. 1991; Hollien et al. 2001; Künzel et al. 1992; Pisoni et al. 1985; Pisoni and Martin 1989; Sobell et al. 1982). However, there was no investigation about whether decrease in speech tempo is firstly influenced by the increase of pause length or by the decrease in the articulation rate. There are more studies that showed that in an alcohol intoxicated state, more speech errors occur (Hollien et al. 2001; Künzel et al. 1992; Trojan and Kryspin-Exner 1968; Zaimov 1969) however, there is also one study (Christenfeld és Creager, 1996) that argues that alcohol reduces the rate of certain speech errors.

Regarding fundamental frequency, results from previous studies are inconsistent. Many of the studies (Behne and Rivera 1990; Hollien et al. 2001; Klingholz et al. 1988; Trojan and

Kryspin-Exner 1968) found increased pitch in an intoxicated state. However, according to Sobell et al. (1982), Johnson et al. (1993) or Gyarmathy (2008), alcohol has no effect or has an ambiguous effect on fundamental frequency: in most of the cases fundamental frequency increases, but in some cases it decreases or does not change. Schiel and Heinrich (2009) found that fundamental frequency typically increased for women and decreased for men. Watanabe et al. (1994) found decreased pitch even for women.

New scientific results

The role of speech pauses in ascribing emotional states in native and in foreign language speech (Study [2])

Thesis 1. Due to changes in only one speech parameter, namely the speech pauses, intensity of the emotion attributed by the listener to the speaker changes. Using other speech parameters as well, listeners automatically ascribe emotional states to the speaker, even in content-neutral speeches.

Thesis 2. The effect of pauses in ascribing emotions can also be detected in German speaking raters while listening to Hungarian speeches. The Hungarian speech samples are rated as sadder, less happy, more scared and less positive when pauses are longer.

Two of our experiments attempted to answer the question how change of one speech parameter, namely the speech pause length influences listeners in ascribing emotional states to the speaker. An other aim of the study was to investigate whether there is a pause length effect when listening to foreign language speech. In our experiment, we used five one-minute long emotionally neutral speech samples. On the recordings, five female volunteers talked about their schools, jobs or an ordinary day in their lives. In these, naturally occurring pauses were manipulated in four different ways: the length of all silent pauses was elongated or shortened in two-two ways.

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. 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. 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.

At the same time, we found differences between the two groups. Austrian participants rated the Hungarian speakers as angrier, sadder, more surprised, more scared, less positive, and more heated than did Hungarian raters. 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.

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.

Furthermore, in previous studies (e.g. Breitenstein et al., 2001), 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 only manipulating pause length and leaving articulation fixed. Finally, the experimental arrangement 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.

The effect of alcohol on speech and on working memory (Studies [3] and [4])

Thesis 3. Alcohol has an effect on disfluencies is speech: taking all types of speech errors (disfluencies, stops, replacements, omissions) into account, subjects make more errors in an alcohol influenced than in a sober state in phonologic, phonetic, morphologic, and syntactic tongue-twisters as well.

Thesis 4. Decrease in speech rate is caused by the increased number of silent pauses and not by the slowdown of the articulation.

Thesis 5. Previous studies produced contradicting results regarding the fundamental frequency change in a moderate alcohol influenced state. According to our results, 0.08% blood alcohol level has no effect on fundamental frequency.

Thesis 6. Alcohol has no effect on the performance on digit span test, but has an effect on the working memory performance needed for the immediate recall of tongue-twisters: in an alcohol influenced state, subjects can recall as many numbers as in a sober state, however it occurs more often in an alcohol influenced state, that subjects are unable to finish the

sentence (stop), they replace the words they can not recall (replacement) or they omit a word (omission).

The third and the fourth study investigated the distal cues: it investigated changes of certain speech features in an alcohol intoxicated state. We also analyzed effects on short term memory. As alcohol has an effect on mood, cognitive functions and motor coordination as well, we assumed that in an alcohol influenced state, speech changes. The task of the participants was to repeat tongue-twisters after consuming 3 times 50 ml of a 40% alcohol concentration shot drink, and in a sober state. We also measured participants' performance on the digit span test both times. As the repeating of tongue-twisters needs capacity of speech production and of memory as well, later, we have separately investigated the effect of alcohol a) on speech features and b) on memory performance. Study 4 includes data about memory performance and detailed analysis of imprecise articulation, while study 3 presents data about fundamental frequency in more detail.

According to the results, in an alcohol influenced state, the number of speech errors and the number of pauses increase; fundamental frequency and articulation rate don't change (Szabó et al., 2012, Tisljár-Szabó et al., 2013). Regarding the effect of alcohol on memory performance, we found that moderate level of alcohol (resulting a 0.08 % blood alcohol level) has no effect on performance on digit span test, but it has an effect on the memory performance that is needed to the repetition of tongue-twisters: in an alcohol influenced state, subjects commit more errors, as there are more stops, replacements and omissions in theirs speech (Szabó et al., 2012).

Discussion

In the PhD thesis I presented studies which investigated how internal states influence speech parameters or how listeners attribute internal states to the speaker based on the co-occurrence of certain speech parameters.

We investigated speech pauses and speech disfluencies that are research topics of conversation analytic and of vocal emotion expression studies as well. Our experiments that investigated the role of speech pauses in ascribing emotions showed that speech pauses play

an important role in what kind of emotions one attributes to the speaker. The same speech samples are perceived to be sadder, less happy, less positive and more scared when pauses are longer. As in everyday situations people usually talk for longer time without interruption, it is highly reasonable to study longer than one sentence utterances. A previous study of us (Szabó, 2008) showed that people in a sad emotional state speak with longer pauses and with a higher pause ration than in a happy emotional state. And, our present study pointed out that speech pauses also function as proximal cues providing information to the listener about the inner state of the speaker.

As it was showed, pause ratio and other speech disfluencies depend on the inner state of the speaker and it also has been demonstrated that pauses can function as proximal cues. As there are still only few studies that analyze 1) sentences in natural situations, 2) longer speech samples, 3) pauses and speech disfluencies depending on different emotional states – further studies are needed on this field. Again, we still don't have exact knowledge about the effect of emotions on cognitive processes needed to speech production.

Since our papers had been written and submitted, several speech technology papers have been published which analyze the possibilities of emotion recognition by human and machine on Hungarian speech samples or by using the ratings of Hungarian listeners (Vicsi 2010, Vicsi et al., 2010, Tóth et al., 2008, Riviello et al. 2012). The novelty of these is that they established emotional speech data bases and they analyze emotion specific prosodic patterns on these. These papers also investigate emotion recognition by machine. They use phrase (an intonation unit) as the basic unit (Vicsi és Sztahó, 2012), in contrast to word level or sentence level units in previous studies. As it is known that beyond the universal cues, language specific effects influence the relationship between emotions and speech, Hungarian studies help to discover similarities and differences in languages.

Beside the role of speech pauses in emotion attribution we investigated the effect of alcohol on speech and on short term memory. It is known that alcohol has an effect on the emotional state, the cognitive and motor processes thus, we assumed that these effects could be tested in repeated tongue-twisters. We analyzed changes within subjects in intoxicated and in sober states. We found that people talk with more speech errors and more pauses in alcohol intoxicated states. The reason of the increased number of errors is the alcohol's effect on the motor-articulation processes. The separated analysis of pauses and articulation rate showed that similarly to previous studies, speech rate decreases in the intoxicated state, however, the decrease in speech rate is due to the increase in the number of pauses and not due to the

decreased articulation rate. According to our results regarding the fundamental frequency, 0.08% blood alcohol level has no effect on fundamental frequency. However, based on previous studies we assume that a higher blood alcohol level would increase fundamental frequency.

We analyzed speech errors in more detail and found that about half of the errors was caused by the imprecise articulation and the other half was caused by the decreased short term memory performance. The fact that moderate level of alcohol (resulting a 0.08 % blood alcohol level) has no effect on performance on digit span test, but has an effect on the memory performance implies that the two tasks (repeating numbers or complex sentences) involves different kind of processes. Interpreting the results in the working memory model (Baddeley, 20003), while the digit span test only uses the phonological loop, repeating complex sentences also requires the performance of the central executives. Thus it is possible that the effect of moderate amount of alcohol only appears in the case of tasks needing the central executives and such an amount of alcohol has no effect on simply repeating. To precise understanding of the process, further studies are needed.

Application, conclusion

The presented papers help us to better understand the relationship between inner states, emotions, personality and speech features. With the help of this knowledge we can let psychologists, doctors, political speech analysts know what kind of changes in speech are to pay attention. Also, the use of the distal and proximal enables to generate emotional speech by machine (e. g. Türk és Schröder, 2008), or to detect emotional (e. g. Vicsi és Sztahó, 2011) or alcohol influenced (e. g. Schiel és Heinrich, 2009) states automatically.

One of the aim of the future research will most likely be to connect researches from different areas and to generate models and computer programs which simultaneously are able to process content, voice, and the interaction structure of the conversation.

List of publication related to the thesis

[1] Tisljár-Szabó, E. (2014) Érzelem és beszéd. In: Pléh Cs. és Lukács Á. (szerk.) A pszicholingvisztika magyar kézikönyve. Budapest, Akadémiai Kiadó. (megjelenés alatt)

[2] Tisljár-Szabó, E., Pléh, C. (2014) Ascribing emotions depending on pause length in native and foreign language speech. *Speech Communication, 56*, 35–48. DOI 10.1016/j.specom.2013.07.009 (IF: 1,283)

[3] Tisljár-Szabó, E., Rossu, R., Varga, V., Pléh, Cs. (2014) The effect of alcohol on speech production. *Journal of Psycholinguistic Research*. (Online first, DOI 10.1007/s10936-013-9278-y) (IF: 0,642)

[4] Szabó, E., Rossu, R., Varga, V., Imréczi, K. (2012) Az alkohol hatása a rövid távú emlékezetre és a beszédprodukciós folyamatokra. *Magyar Pszichológiai Szemle, 67*, 3, 409-430. DOI: 10.1556/MPSzle.67.2012.3.1.

Conferences

Szabó, E. (2009) A néma szünetek időtartamának hatása az érzelmi állapot észlelésére. *Magyar Számítógépes Nyelvészeti Konferencia,* Szeged. (poszter)

Szabó, E., Imréczi, K., Rossu, R., Varga, V. (2010) Az alkohol hatása a munkamemóriára és a beszédprodukciós folyamatokra. *Magyar Pszichológiai Társaság Nagygyűlése – Egyén és Kultúra,*Pécs. (poszter)

Szabó, E. (2011) Érzelemfelismerés idegen nyelvű beszédben. *Magyar Pszichológiai Társaság Nagygyűlése,* Budapest. (poszter)

Szabó, E. (2011) Ascribing emotions in foreign language speech. *4th Joint Action Meeting*, Vienna, Austria. (poszter)

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): p. 481-490.

Baddeley, A. (2003) Working memory: looking back and looking forward. *Neuroscience*, 4, 10, 829–839.

Baum-Baicker, C. (1985) The psychological benefits of moderate alcohol consumption: A review of the literature. *Drug and Alcohol Dependence*, 15, 4, 305–322.

Behne, D. M., & Rivera, S. M. (1990) Effects of alcohol on speech: Acoustic analysis of spondees. *Research on Speech Perception,* 16, 263–291.

Behne, D. M., Rivera, S. M., Pisoni, D. B. (1991) Effects of alcohol on speech. *Journal of the Acoustical Society of America,* 90, 4, 2311-2311.

Behne, D. M., Rivera, S. M., Pisoni, D. B. (1991) Effects of alcohol on speech. *Journal of the Acoustical Society of America,* 90, 4, 2311-2311.

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]. *Z Exp Angew Psychol,*. **35**(2): p. 167-200.

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

Burkhardt, F. and W.F. Sendlmeier (2000) Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis, in *ISCA Workshop (ITRW) on Speech and Emotion2000*: Belfast.

Cahn, J. E. (1990) The Generation of Affect in Synthesized Speech. *Journal of the American Voice I/O Society,* 1-19.

Carlson, R., (1992) Synthesis - Modeling Variability and Constraints. *Speech Communication, 11*(2-3): p. 159-166.

Christenfeld, N., Creager, B. (1996) Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology,* 70, 3, 451-460.

Deppermann, A. és 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.

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

Grandjean, D., Bänziger, T., Scherer, K. R. (2006) Intonation as an interface between language and affect. *Progress in Brain Research, 156*, 235–268.

Hofmann, S.G., Gerlach, A.L., Wender, A., Roth, W.T. (1997) Speech disturbances and gaze behavior during public speaking in subtypes of social phobia. *J Anxiety Disord, 11*(6), 573-85.

Hollien, H., Dejong, G., Martin, C. A., Schwartz, R., Liljegren, K. (2001) Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America,* 110, 3198-206.

Johnson, K., Southwood, M. H., Schmidt, A. M., Mouli, C. M., Holmes, A. T., Armstrong, A. A., . . . WiIson, A. S. (1993). *A physiological study of the effects of alcohol on speech and voice.* Paper presented at the 22nd annual Symposium on the Care of the Professional Voice at the Voice Foundation.

Jones, M. K., Jones, B. M. (1980) The Relationship of Age and Drinking Habits to the Effects of Alcohol on Memory in Women. *Journal of Studies on Alcohol,* 41, 1, 179–186.

Künzel, H. J., Braun, A., Eyshold, U. (1992) *Einfluss von Alkohol auf Sprache und Stimme*. Kriminalistik Verlag, Heidelberg

Ladd, D.R., Silverman, K., Tolkmitt, F., Bergmann, G., Scherer, K.R. (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., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., Michelgård, Å., Appel, L., Fredrikson, M., Furmark, T. (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. *J Abnorm Psychol, 53*(1), 1-15.

Maylor, E. A., Rabbitt P. M. A (1987) Effect of alcohol on rate of forgetting. *Psychopharmacology,* 91, 2, 230–235.

Miller, L. L., Dolan, M. P. (1974) Effects of alcohol on short term memory as measured by a guessing technique. *Psychopharmacology,* 35, 4, 353–364.

Nordby, K., Watten, R. G., Raanaas, R. K., Magnussen, S. (1999) Effects of Moderate Doses of Alcohol on Immediate Recall of Numbers: Some Implications for Information Technology. *Journal of Studies on Alcohol*, 60, 6, 873–878.

Pell, M.D., Skorup, V. (2008) Implicit processing of emotional prosody in a foreign versus native language. Speech Communication, 50(6), 519-530.

Pisoni D. B., Martin C. S. (1989) Effects of alcohol on the acoustic-phonetic properties of speech: perceptual and acoustic analyses. *Alcoholism: Clinical Experimental Research,* 13, 4, 577–587.

Pisoni, D. B., Hathaway, S. N., Yuchtman, M. (1985) *Effects of alcohol on the acoustic-phonetic properties ofspeech: Finalreport to* GM *Research Laboratories.* (SRL Technical Note No. 85-03). Bloomington: Indiana University.

Pope, B., Blass, T., Siegman, A.W., Raher, J. (1970) Anxiety and depression in speech. *J Consult Clin Psychol,. 35*(1), 128-33.

Riviello, M. T., Esposito, A., Vicsi, K (2012) A Cross-Cultural Study on the Perception of Emotions: How Hungarian Subject Evaulate American and Italian Emotional Expressions. *Lecture Notes In Computer Science* 7403, pp. 424-433.

Scherer, K. R. (2003). Vocal communication of emotion: A rewiev of research paradigms. *Speech Communication 40*. 227–256.

Scherer K. R. (2013) *Vocal markers of emotion*: Comparing induction and acting elicitation. Computer Speech & Language 27(1): 40-58

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.

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.

Schiel, F., Heinrich, C. (2009) Laying the Foundation for In-car Alcohol Detection by Speech. *Proceedings of the INTERSPEECH 2009,* Brighton, UK, pp. 983–986.

Sobell, L. C., Sobell, M. B., & Coleman, R. F. (1982) Alcohol-induced disfluency in nonalcoholics. *Folia Phoniatrica*, 34, 316–323.

Sobell, L. C., Sobell, M. B., & Coleman, R. F. (1982) Alcohol-induced disfluency in nonalcoholics. *Folia Phoniatrica*, 34, 316–323.

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.

Trojan, F., Kryspin-Exner, K. (1968) The decay of articulation under the influence of alcohol and paraldehyde. *Folia Phoniatrica (Basel),* 20, 4, 217-238.

Tóth, Sz. L., Sztahó, D., Vicsi, K. (2008) Speech Emotion Perception by Human and Machine. In: Proceeding of COST Action 2102 International Conference: Revised Papers in Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. Springer, pp. 213-224.

Türk, O., Schröder, M. (2008) A comparison of voice conversion methods for transforming voice quality in emotional speech synthesis. *Proc. Interspeech 2008*, Brisbane, Australia.

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.

Vicsi K. (2010) Érzelemfelismerés. In: Németh G, Olaszy G (szerk.) *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Budapest, Akadémiai Kiadó, pp. 392-401.

Vicsi, K., Sztahó, D. (2011) Problems of the Automatic Emotion Recognitions in Spontaneous Speech; An Example for the Recognition in a Dispatcher Center. Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues. Lecture Notes in Computer Science Volume 6456, pp 331-339.

Vicsi, K, Sztahó, D (2012) Recognition of emotions on the basis of different levels of speech segments. Journal Of Advanced Computational Intelligence And Intelligent Informatics 16, 2, pp. 335-340.

Vicsi, K., Sztahó D., Kiss, G., Czira, A. (2010) Spontán beszédben rejlő nem verbális hangjelenségek - érzelmek, hanggesztusok - vizsgálata. In: Tanács A., Vincze V. (szerk.) MSZNY 2010: VII: Magyar Számítógépes Nyelvészeti Konferencia. Szeged,

Magyarország, 2010.12.02-2010.12.03. Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, pp. 249-260.

Watanabe, H., Shin, T., Matsuo, H., Okuno, F., Tsuji, T., Matsuoka, M., . . . Matsunaga, H. (1994) Studies on vocal fold injection and changes in pitch associated with alcohol intake. Journal of Voice, 8, 4, 340-346.

Zaimov, K. (1969) Die Sprachstörungen als Kriterium der Bewusstseinsstörungen. *Psychiatrie, Neurologie, und Medizinische Psychologie,* 21, 218–225.