Research Article

Heart Rate Responses to Synthesized Affective Spoken Words

Mirja Ilves and Veikko Surakka

*Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer-Human Interaction (TAUCHI),
School of Information Sciences, University of Tampere, Kanslerinrinne 1, FI-33014 University of Tampere, Finland*

Correspondence should be addressed to Mirja Ilves, mirja.ilves@sis.uta.fi

Received 7 March 2012; Revised 20 June 2012; Accepted 5 July 2012

Academic Editor: Eva Cerezo

Copyright © 2012 M. Ilves and V. Surakka. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present study investigated the effects of brief synthesized spoken words with emotional content on the ratings of emotions and heart rate responses. Twenty participants' heart rate functioning was measured while they listened to a set of emotionally negative, neutral, and positive words produced by speech synthesizers. At the end of the experiment, ratings of emotional experiences were also collected. The results showed that the ratings of the words were in accordance with their valence. Heart rate deceleration was significantly the strongest and most prolonged to the negative stimuli. The findings are the first suggesting that brief spoken emotionally toned words evoke a similar heart rate response pattern found earlier for more sustained emotional stimuli.

1. Introduction

Verbal communication is unique to humans, and speech is an especially effective means to communicate ideas and emotions to other people [1]. McGregor [2] argued that spoken language is a more primary and more fundamental means of communication than written language. In speech, both verbal meaning and prosodic cues within the speech can communicate emotions; however, there is little research on the role of the verbal meaning of spoken words to human emotions. Although studies concerning the emotional processing of the verbal content of speech are rare, the scope of emotion studies has recently broadened from studying the reactions to the pictures of emotional scenes and human faces to visually presented linguistic stimuli. In a way, emotionally charged spoken stimuli uttered in a monotone or a neutral tone of voice partly parallels written text. By this, it is meant that only lexical contents of the stimuli offer knowledge about emotion, so the results about visually presented written words can provide some background references for studying reactions to spoken emotional words.

Studies using event related potential (ERP) measurements have repeatedly found that early cortical responses to visually presented words with emotional content are enhanced as compared to ERPs to neutral words. This

suggests that the emotional content of a word is identified at an early lexical stage of processing (e.g., [3–5]). In addition, there is evidence that written emotionally negative words evoke larger activation of *corrugator supercilii* (i.e., frowning) facial muscle than positive words do [6–8]. There are also studies that have found larger startle reflex to unpleasant words than to neutral and positive words during shallow word processing [9, 10]. Further, there is some evidence that written words evoke autonomic nervous system responses. One study showed that heart rate decelerated more in response to unpleasant words than to other word types [11]. Based on these studies, physiological response patterns to emotional words seem to be quite consistent with the earlier findings about the reactions to emotional pictures and sounds.

Of course, the results of the responses to visually presented words are not directly applicable to spoken words, and more research is to be done to identify the role of the verbal content of spoken language in human emotional processing. In order to study purely the effects of the verbal content of spoken words to human emotion system, speech synthesizers offer good opportunities; this is because they offer good controllability over timing and prosodic cues related to the nonverbal expressions of emotions like loudness and the range of the fundamental frequency of voice. Controlling

out the variation in nonverbal cues offers the possibility to study more purely the effects of the lexical meaning of spoken words. This results in a monotone tone of voice; however, there is evidence that such stimuli work well in the context of emotion research. A recent study investigated how spoken emotional stimuli pronounced by a monotone tone of synthesized voice activate the human emotional system. The findings showed that the verbal content of spoken sentences evoked emotion-relevant subjective and pupil responses in listeners [12]. Naturally, the knowledge about the effects of synthetically produced messages is also important in an application point of view; this is because interfaces that utilize speech synthesis have become increasingly popular.

Based on earlier research, heart rate measurement could provide new and valuable knowledge about the processing of spoken words; this is because previous studies have suggested that heart rate is sensitive to both cognitive (i.e., attention) and emotional aspects of stimulus processing. These studies have utilized the dimensional theory of emotions [13]. In earlier research, the most commonly used dimensions have been valence, reflecting the pleasantness of an emotional state, and arousal, which reflects the level of emotional activation. Lang et al. [14] have suggested that these dimensions relate to the functioning of the human motivational system, which guides us to approach or withdraw from objects. There is evidence that heart rate accelerates when people imagine emotional material [14] and decelerates during the perception of emotional information, such that the deceleration is largest and most prolonged to negative stimuli [15].

Based on these types of findings, Lang et al. [16] have suggested a special model known as the defense cascade model. This model suggests that a new stimulus in an environment causes an orienting response accompanied with a brief heart rate deceleration. Sustained heart rate deceleration relates to the increased and continued allocation of attentional resources towards somehow-threatening stimulus, which, however, does not require immediate action. If the stimulus is nonthreatening, the heart rate starts to recover back to the prestimulus onset level. Lang et al. [14] suggested that although the connection between valence and heart rate is evident, it is modest. Codispoti et al. [17] also suggested that a decelerating pattern of heart rate in visual contexts presumes sustained sensory stimulation. Research concerning heart rate responses to spoken stimuli can provide important knowledge about the meaning of speech, particularly verbal content, in emotional processing. It is noteworthy that the processing of emotion-related auditory information can have a significant role for human behavior because the hearing of vocally uttered sounds may have had a significant meaning for survival in many situations. Thus, the present study investigated the effects of brief emotionally negative, neutral, and positive synthesized spoken messages to heart rate and the ratings of emotions.

2. Method

2.1. Participants. Thirty-one volunteer students participated in the experiment. Data from eight participants were

discarded from the analysis because of poor signal-to-noise ratio to guarantee the high quality of the analyzed heart rate data. In addition, data from three participants had to be rejected due to other equipment failures. Thus, data from 20 participants were used for the analysis (11 males, M_{age} : 27.8 years, range: 19–51 years). The participants were native speakers of Finnish and had normal hearing by their own report. All participants signed a written consent.

2.2. Equipment. The presentation of stimuli and rating scales were controlled by E-Prime [18] experiment generator software. The stimuli were audible via two loudspeakers placed in front of the participant. Continuous ballistocardiographic heart rate responses (beats/min) were measured with sampling rate of 500 Hz using the EMFi chair. The Quatech DAQP-16 card digitized the heart rate signal to a PC with a Windows XP operating system. The EMFi chair is a regular-looking office chair that has been developed and tested for unobtrusive measuring of heart rate changes [19–21]. The chair is embedded with electromechanical film (EMFi) sensors in its seat, backrest, and armrests. EMFi is a low-cost, cellular, electrically charged polypropylene film that senses changes in pressure and, thus, can be used to measure ballistocardiographic heart rate.

2.3. Stimuli. Stimuli consisted of 15 different affective words selected from the ANEW [22] based on their mean ratings in valence and arousal scales (ANEW stimuli numbers: 8, 37, 46, 227, 251, 266, 305, 494, 591, 613, 614, 759, 974, 1001, and 1015). Five stimuli were strongly positive and arousing, five were strongly negative and arousing, and five were neutral words. The words were translated in Finnish, and their lengths were matched across the categories.

The male voice of three Finnish speech synthesizers produced all the words. The used synthesis techniques were formant synthesis [23], diphone synthesis [24], and unit selection [25]. Each word was repeated three times, resulting in 45 stimulus words in total. Fundamental frequency of the voices was set to 100 Hz. The volume normalization was set at 75 dB using Wave Surfer 1.8.5 program, and the changes in intonation were set to zero, resulting in a monotone tone of voice.

2.4. Experimental Procedure. First, the electrically shielded and sound attenuated laboratory was introduced to the participant, and then the participant sat in the EMFi chair. The experimenter informed the participant that the aim was to study reactions to auditory stimuli and that the experiment would consist of a listening and a rating phase. The participant was to relax, sit still, and concentrate on listening to the set of words produced by speech synthesis. Then, in addition to the experimenter, speech synthesizer gave the instructions regarding the experimental task in order to get participant used to the monotone voice of the speech synthesizer. This synthesizer was not among the synthesizers used in the actual experiment. After giving the instructions, the experimenter left the room, and the stimulus presentation begun. After the experiment, the EMFi

chair and the purpose of using it was explained to each participant.

Interstimulus interval was 15 seconds. A fixation-cross was in view five seconds before and after each stimulus presentation. After the fixation-cross disappeared, the question “Did you understand the word?” and the answer options “no” and “yes” appeared at the screen. The pressing of the left button of a response box indicated that she/he did not understand the word and the right button that she/he understood the word. The order of the stimulus presentation was fully randomized.

After the experimental phase, the participant heard the words again and rated her/his emotional experience during each stimulus on three dimensions: valence, arousal, and approachability. The ratings were given on nine-point bipolar scales. The valence scale ranged from unpleasant to pleasant, the arousal scale from calm to aroused, and the approachability scale from withdrawable to approachable. The center of all the scales represented neutral experience. Before the rating session, the scales were explained to the participant, and the giving of the ratings was rehearsed through two exercise stimuli. The stimuli were presented randomly, rating scales were presented on the display, and a keyboard was used to give the ratings. Finally, the participant was debriefed about the purpose of the study. None of the participants was aware that the chair was used to measure heart rate responses until after the experiment.

2.5. Data Analysis. First, artifacts caused, for example, by body movements were removed from the heart rate data with the algorithm described in Anttonen et al. [21]. Then, the data were baseline corrected using a one-second pre-stimulus baseline. The stimuli that included $\geq 50\%$ artifacts during the baseline period or during five seconds period from stimulus offset were discarded from further analysis (i.e., 19% of the data). Then, mean heart rate values both for the overall data and for second-by-second data five seconds from the stimulus offset were calculated. Finally, the data were categorized according to the stimulus categories.

The mean heart rate responses five seconds from the stimulus offset did not differ significantly between the stimuli that were understood and the stimuli that were not understood $t(19) = 1.550$, ns. Thus, the words that were judged difficult to understand were included in the analysis. It is likely that at least part of the words that participant judged as unclear were actually heard and understood, but he/she answered no when the word sounded dim. Further, there were no differences between the understandability of the negative and positive words $t(19) = 1.901$, ns. Instead, the neutral words were perceived as significantly more understandable than the negative $t(19) = 4.717$, $P < .001$, or positive $t(19) = 3.823$, $P < .01$ words. In addition, the heart rate responses did not differ significantly between the men and women $t(18) = 0.496$, ns.

The data were analyzed with repeated measures ANOVAs with Greenhouse-Geisser adjusted degrees of freedom when necessary. Multiple post hoc pairwise comparisons were Bonferroni corrected.

3. Results

3.1. Heart Rate. Figure 1 shows the mean heart rate responses, averaged over all synthesizers, to the different emotion categories during the period of five seconds after the stimulus offset. The left side shows grand averages during the period of 5 seconds from the stimulus offset. The right side shows responses on a second-by-second basis, both during the stimulation and five seconds from the stimulus offset. Figure 1 clearly shows that the heart rate responses to the different stimulus categories were diverse. Heart rate response seemed to be larger and more prolonged to the negative stimuli than to the positive or neutral stimuli. First, we tested the effect of the synthesizer on the heart rate responses. One-way ANOVA did not reveal a significant effect of the synthesizer to heart rate, $F(2,38) = 1.09$, ns.

One-sample t -test revealed a statistically significant deceleration from prestimulus baseline for the negative ($t(19) = 3.32$, $P < .01$, $d = .74$) and neutral ($t(19) = 2.40$, $P < .05$, $d = .54$) stimuli. Decrease after the positive stimuli was not statistically significant ($t(19) = 0.90$, $P > .05$). Although one-way ANOVA with emotion content as a within subject factor was only nearly significant ($F(2, 38) = 3.12$, $P = .056$, $\eta^2 = .05$), the linear trend between emotion categories was significant ($F(1, 19) = 4.95$, $P < .05$). Thus, the deceleration was largest to the negative words.

The second-by-second analysis revealed a statistically significant effect of emotion category at the fifth second from the stimulus offset ($F(2, 38) = 4.80$, $P < .05$, $\eta^2 = .07$); negative versus positive ($t(19) = 2.64$, $P < .05$). Difference between the negative and neutral words or between the neutral and positive words was not significant. At that time point, the linear trend was also statistically significant ($F(1, 19) = 6.97$, $P < .05$).

3.2. Subjective Ratings. Table 1 shows the results of the analysis of the subjective ratings. The statistical analysis showed a significant effect for the ratings of valence. Post hoc pairwise comparisons were all statistically significant. The positive words ($M = 5.71$) evoked significantly more positive ratings of valence than the negative ($M = 3.75$) or neutral words ($M = 4.89$) did. The negative words evoked significantly more negative ratings of valence than the neutral words did. There was also a significant effect in the ratings of arousal. The negative words ($M = 5.95$) were significantly more arousing than the neutral ($M = 5.21$) or positive words ($M = 5.08$). Further, there was a significant effect in the ratings of approachability. The positive words ($M = 5.47$) were rated as significantly more approachable than the negative ($M = 3.44$) or neutral words ($M = 4.82$). The neutral words were rated as more approachable than the negative words.

4. Discussion

This study was the first showing significant heart rate changes to brief spoken emotionally toned words. The findings are of particular concern to responses to synthetically produced words. Heart rate deceleration five seconds from the stimulus

FIGURE 1: (a) Averaged heart rate responses and SEM to the negative, neutral, and positive words during the period of five seconds after the stimulus offset. (b) Averaged second-by-second heart rate changes by stimulus categories during the stimulus and five seconds after the stimulus offset.

TABLE 1: Statistical results of the ratings of valence, arousal, and approachability.

Rating scale	ANOVA	Effect size η^2	Pairwise comparisons
Valence	$F(2, 38) = 30.60, P < .001$	$\eta^2 = .54$	Negative < neutral $t(19) = 5.99, P < .001$ Negative < positive $t(19) = 5.67, P < .001$ Neutral < positive $t(19) = 4.44, P < .001$
Arousal	$F(2, 38) = 13.57, P < .001$	$\eta^2 = .19$	Negative > neutral $t(19) = 4.59, P < .001$ Negative > positive $t(19) = 4.30, P < .01$ Neutral versus positive $t(19) = 0.76, ns$
Approachability	$F(1, 23) = 29.30, P < .001$	$\eta^2 = .48$	Negative < neutral $t(19) = 7.94, P < .001$ Negative < positive $t(19) = 5.67, P < .001$ Neutral < positive $t(19) = 2.64, P < .05$

offset was largest to the negative stimuli. Further, the deceleration was the most prolonged to the negative stimuli, such that the difference between heart rate responses to the negative and positive stimuli was statistically significant at the fifth second from the stimulus offset. This was because at the fifth second from the offset of the positive stimulation heart rate had recovered nearly back to the baseline, while heart rate response to the negative stimulation remained decelerated. This finding is in line with the earlier studies that have investigated heart rate changes second-by-second (e.g., [19]). Many studies (e.g., [15, 17, 26]) have used other types of analysis, like analyzing only the averages over several seconds of stimulations, which do not reveal the second-by-second functioning of the heart. Overall, the current results showed that heart rate changes followed a triphasic form, which has been a typical pattern of response during

the perception of emotional stimulation [16]. This form is characterized by an initial deceleration, then an acceleration following a late deceleration so that heart rate deceleration is altogether stronger during negative than during positive stimulation. The ratings of the stimuli were in accordance with negative, neutral, and positive emotion categories.

Although the earlier findings of responses to emotional stimuli have been mainly consistent, Codispoti et al. [17] suggested that heart rate responses to briefly presented affective pictures are different from responses to longer stimuli. In their study, initial heart rate deceleration six seconds from the stimulus onset was minimal, and picture valence had no significant effect on heart rate responses. Thus, they concluded that a clear decelerating response to affective stimuli in visual contexts presumes more sustained sensory stimulation. In contrast to this, the present results suggest

that the heart reacts differently to the spoken verbal material and that even very brief spoken words with lexical emotional content seem to evoke a corresponding decelerated heart rate response pattern as found earlier with longer emotional stimuli. Although there were some difficulties to understand some of the words, there were no significant differences between the understandability of the negative and positive words. Thus, the different heart rate responses to the negative and positive words reflect the differences in the emotional valence of the words, not the pronunciation quality of the word.

The current findings could be explained by differences in the processing of visual and auditory stimuli or in the processing of verbal and pictorial material. It is known that different senses process sensory information in different ways. For example, there are different sensory stores for each sense. There is evidence that auditory sensory information is available much longer (i.e., seconds) in a sensory storage than visual information (i.e., only a few hundred milliseconds). It has been proposed that auditory sensory store consists of two phases [27, 28]. In the initial phase, the stimulus is unanalyzed and it is stored for several hundred milliseconds. In the second phase, the information is stored for a few seconds and the content is partly analyzed. This means that while the auditory information is still available in the sensory store, the visual information has passed to the conscious processing of the working memory and has faded from the sensory store. Thus, it is possible that the hearing of brief lexical stimuli evokes similar heart rate responses to those to more continuous visual sensory stimulation.

Secondly, there is evidence that the processing of pictures and words activates the same brain areas [29]. However, studies with ERPs [30] and categorizing tasks [31] have shown that the processing time of words is slower than the processing time of pictures. This difference has been found by comparing the processing times between the visually presented words and pictures. Further, there is evidence that the processing of speech takes even a little longer than reading (see [32]). These differences may reflect the case that the processing of words and pictures is different. Following this, their effects seem to differ from each other. Because the processing time of spoken words is slower than the processing time of pictures, it may be also that listening binds the attentional resources longer than watching pictures does.

There is also some debate as to what extent the processing of emotional words affects physiological activation. Mainly, the findings about physiological responses to emotional words have been consistent with the earlier findings about the reactions to emotional pictures and sounds. However, it has also been suggested that the emotional content of written words does not automatically result in autonomic activation [33] or that the physiological responses are much smaller for words than for emotional pictures and sounds (e.g., [8]). Instead, the current (as well as some previous [12]) results suggest that perceiving the emotional content of the spoken words induces emotional reactions in the listeners even though the words were produced by the monotone voice of speech synthesis.

This may reflect the importance of speech and listening for humans. First, hearing has some advantages in comparison with other senses. For example, vision is limited to the position of the eyes while hearing is relatively independent from the position of the ears and the head. Thus, hearing is a primary system for receiving warnings from others [34]; consequently, people may be highly responsive to vocal emotional messages. Second, speech has a central role when people communicate each other and build social relationships [1]. Thus, humans have evolved to receive very effectively messages delivered through speech; perhaps for this reason the emotional words produced even by the flat, synthesized voice cause emotional responses in humans. In future research, it would be interesting to study whether there are any differences in emotional responses to synthesized and natural voices. There are some findings showing that increasing the human likeness of synthesized voice strengthens the emotional responses to voice messages [12, 35, 36]. On the other hand, the used synthesis technique did not affect heart rate responses in the present study. Thus, although it is possible that the emotional responses to human speech would be stronger as compared to synthetic voice, the present results together with the previous findings suggest that synthetic speech with emotional content also evokes emotional responses in humans.

In conclusion, the present results are the first to suggest that brief spoken words with emotional content can evoke the decelerated heart rate response pattern similar to those previously obtained only with longer stimuli. Further, the affective words evoked emotion-related responses in people, even though the words were produced by the artificial-sounding monotonous synthesized voices. Thus, it seems that the emotional expressions produced by a computer can evoke emotional responses in a human similar to responses to the emotional expressions of another human. The knowledge that speech-based computer systems can induce emotions in people is important for the field of human-computer interaction when creating and designing computerized interactive environments and devices.

Acknowledgments

The authors would like to thank all the participants of the study. This research was supported by the Doctoral Program in User-Centered Information Technology (UCIT) and a grant from the University of Tampere.

References

- [1] G. A. de Laguna, *Speech: Its Function and Development*, Indiana University Press, Bloomington, Ind, USA, 1963.
- [2] W. McGregor, *Linguistics: An Introduction*, Continuum International, New York, NY, USA, 2009.
- [3] C. Herbert, M. Junghofer, and J. Kissler, "Event related potentials to emotional adjectives during reading," *Psychophysiology*, vol. 45, no. 3, pp. 487–498, 2008.
- [4] J. Kissler, C. Herbert, P. Peyk, and M. Junghofer, "Buzzwords: early cortical responses to emotional words during reading:

- research report," *Psychological Science*, vol. 18, no. 6, pp. 475–480, 2007.
- [5] G. G. Scott, P. J. O'Donnell, H. Leuthold, and S. C. Sereno, "Early emotion word processing: evidence from event-related potentials," *Biological Psychology*, vol. 80, no. 1, pp. 95–104, 2009.
 - [6] M. Bayer, W. Sommer, and A. Schacht, "Reading emotional words within sentences: the impact of arousal and valence on event-related potentials," *International Journal of Psychophysiology*, vol. 78, no. 3, pp. 299–307, 2010.
 - [7] F. Foroni and G. R. Semin, "Language that puts you in touch with your bodily feelings: the multimodal responsiveness of affective expressions," *Psychological Science*, vol. 20, no. 8, pp. 974–980, 2009.
 - [8] J. T. Larsen, C. J. Norris, and J. T. Cacioppo, "Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii," *Psychophysiology*, vol. 40, no. 5, pp. 776–785, 2003.
 - [9] C. Herbert, J. Kissler, M. Junghöfer, P. Peyk, and B. Rockstroh, "Processing of emotional adjectives: evidence from startle EMG and ERPs," *Psychophysiology*, vol. 43, no. 2, pp. 197–206, 2006.
 - [10] C. Herbert and J. Kissler, "Motivational priming and processing interrupt: startle reflex modulation during shallow and deep processing of emotional words," *International Journal of Psychophysiology*, vol. 76, no. 2, pp. 64–71, 2010.
 - [11] T. W. Buchanan, J. A. Etzel, R. Adolphs, and D. Tranel, "The influence of autonomic arousal and semantic relatedness on memory for emotional words," *International Journal of Psychophysiology*, vol. 61, no. 1, pp. 26–33, 2006.
 - [12] M. Ilves and V. Surakka, "Emotions, anthropomorphism of speech synthesis, and psychophysiology," in *Emotions in the Human Voice*, K. Izdebski, Ed., Culture and Perception, pp. 137–152, Plural, San Diego, Calif, USA, 2009.
 - [13] M. M. Bradley, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
 - [14] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, attention, and the startle reflex," *Psychological Review*, vol. 97, no. 3, pp. 377–395, 1990.
 - [15] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.
 - [16] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Motivated attention: affect, activation, and action," in *Attention and Orienting—Sensory and Motivational Processes*, P. J. Lang, R. F. Simons, and M. T. Balaban, Eds., pp. 97–135, Erlbaum, Mahwah, NJ, USA, 1997.
 - [17] M. Codispoti, M. M. Bradley, and P. J. Lang, "Affective reactions to briefly presented pictures," *Psychophysiology*, vol. 38, no. 3, pp. 474–478, 2001.
 - [18] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime User's Guide*, Psychology Software Tools, Pittsburgh, Pa, USA, 2002.
 - [19] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pp. 491–499, April 2005.
 - [20] J. Anttonen and V. Surakka, "Music, heart rate, and emotions in the context of stimulating technologies," in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII '07)*, pp. 290–301.
 - [21] J. Anttonen, V. Surakka, and M. Koivuluoma, "Ballistocardiographic responses to dynamic facial displays of emotion while sitting on the EMFi chair," *Journal of Media Psychology*, vol. 21, no. 2, pp. 69–84, 2009.
 - [22] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): stimuli, instruction manual and affective ratings," Tech. Rep. C-1, University of Florida, Gainesville, Fla, USA, 1999.
 - [23] T. Saarni, *Segmental Durations of Speech [Doctoral Dissertation]*, University of Turku, Finland, 2010.
 - [24] Suopuhe [speech synthesizer], <http://www.ling.helsinki.fi/suopuhe/english.shtml>.
 - [25] Bitlips [speech synthesizer], <http://www.bitlips.fi/index.en.html>.
 - [26] O. Pollatos, B. M. Herbert, E. Matthias, and R. Schandry, "Heart rate response after emotional picture presentation is modulated by interoceptive awareness," *International Journal of Psychophysiology*, vol. 63, no. 1, pp. 117–124, 2007.
 - [27] N. Cowan, "On short and long auditory stores," *Psychological Bulletin*, vol. 96, no. 2, pp. 341–370, 1984.
 - [28] N. Cowan, "Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system," *Psychological Bulletin*, vol. 104, no. 2, pp. 163–191, 1988.
 - [29] P. Bright, H. Moss, and L. K. Tyler, "Unitary vs multiple semantics: PET studies of word and picture processing," *Brain and Language*, vol. 89, no. 3, pp. 417–432, 2004.
 - [30] A. Schacht and W. Sommer, "Emotions in word and face processing: early and late cortical responses," *Brain and Cognition*, vol. 69, no. 3, pp. 538–550, 2009.
 - [31] A. Roelofs, "Dynamic of the attentional control of word retrieval: analyses of response time distributions," *Journal of Experimental Psychology*, vol. 137, no. 2, pp. 303–323, 2008.
 - [32] K. Rayner and C. Clifton, "Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research," *Biological Psychology*, vol. 80, no. 1, pp. 4–9, 2009.
 - [33] M. Bayer, W. Sommer, and A. Schacht, "Emotional words impact the mind but not the body: evidence from pupillary responses," *Psychophysiology*, vol. 48, no. 11, pp. 1553–1561, 2011.
 - [34] B. Scharf, "Auditory attention: the psychoacoustical approach," in *Attention*, H. Pashler, Ed., pp. 75–117, Psychology Press, Hove, UK, 1998.
 - [35] M. Ilves and V. Surakka, "Subjective and physiological responses to emotional content of synthesized speech," in *Proceedings of the International Conference on Computer Animation and Social Agents (CASA '04)*, N. Magnenat-Thalmann, C. Joslin, and H. Kim, Eds., pp. 19–26, Computer Graphics Society, Geneva, Switzerland, 2004.
 - [36] M. Ilves, V. Surakka, T. Vanhala et al., "The effects of emotionally worded synthesized speech on the ratings of emotions and voice quality," in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII '11)*, Part I, S. D'Mello et al., Ed., vol. 6974 of *Lecture Notes in Computer Science*, pp. 588–598, Springer, 2011.