Language specific effects of emotion on phoneme duration

Martijn Goudbeek[1], *Mirjam Broersma*[2]

[1]University of Tilburg, The Netherlands
[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
m.b.goudbeek@uvt.nl, mirjam@mirjambroersma.nl

Abstract

This paper presents an analysis of phoneme durations of emotional speech in two languages: Dutch and Korean. The analyzed corpus of emotional speech has been specifically developed for the purpose of cross-linguistic comparison, and is more balanced than any similar corpus available so far: a) it contains expressions by both Dutch and Korean actors and is based on judgments by both Dutch and Korean listeners; b) the same elicitation technique and recording procedure were used for recordings of both languages; and c) the phonetics of the carrier phrase were constructed to be permissible in both languages. The carefully controlled phonetic content of the carrier phrase allows for analysis of the role of specific phonetic features, such as phoneme duration, in emotional expression in Dutch and Korean. In this study the mutual effect of language and emotion on phoneme duration is presented.

Index Terms: emotional speech, phonetic analysis, duration, cross linguistic comparison.

1. Introduction

In this paper, we present durational measures of phonetic segments in a corpus of Dutch and Korean vocal emotion expressions: the Demo (Dutch Emotion) / Kemo (Korean Emotion) corpus. In contrast to existing corpora of vocal emotion expressions, the present corpus has been specifically developed for the purpose of cross-linguistic and cross-cultural comparison. Therefore, it is more balanced than any materials available so far. The corpus contains a comparatively large number of emotions (eight) uttered by a large number of speakers (eight Dutch speakers and eight Korean speakers). Further, the phonetic content of the expressions has been carefully selected to enable the analysis of the role of specific phonetic features in the expression and recognition of emotion in the two languages.

Basic emotions like joy, anger, fear, and sadness have been shown to be recognized well above chance levels between cultures [1]. The increasing body of evidence that shows recognition of emotional expressions from another culture to be above chance and shows certain culturally invariant properties in the expression of emotion, has been taken as support for "basic emotion theory"[2, 3]. The line of reasoning is that if members from a radically different culture are able to understand which emotion is being expressed and express these emotion in more or less similar ways, then this expression can be said not to depend on cultural factors, but must be universal.

Most of the research on cross-cultural factors on the expression of emotion investigates facial expressions, although some notable efforts have been made in the past with respect to vocal expression [4, 5]. Recently, the effects of language and culture on the vocal encoding and decoding of emotion has been the topic of many studies [6, 7, 8, 9, 10]. In these studies, the emotions are usually expressed in meaningless phrases or nonverbal affect vocalizations. Both contain a minimum of semantic information while remaining linguistically valid. Nonverbal affect vocalizations have the advantage over meaningless phrases that they are more natural. However, they have the serious disadvantage of possibly being semantically loaded (e.g., "yuck" might be a universal expression of disgust, regardless of its nonverbal realization). These studies showed that vocal expressions of (basic) emotions can be accurately decoded when listening to expressions in a foreign language. Recent work by Sauter and colleagues [8] showed that even judges of the isolated Himba community manage to accurately decode emotional vocalizations of British speakers. This suggests that it is certainly likely that some aspects of vocal emotion expression are present in most if not all languages.

Nevertheless, many studies, including [6] and [8] also stress that there are language specific elements in the communication of vocal emotion. Pell et al. label this an "in-group" advantage (participants perform better in their native language than in a nonnative language), and Scherer et al. formulate the bolder "language distance hypothesis": based on their finding that listeners of a closely related language like English were better at categorizing German emotional expressions than listeners of a more remote language like Indonesian, Scherer et al. conclude that linguistic similarity plays an important role when decoding emotional expressions in a foreign language.

Taken together, these studies strongly suggest that expression and understanding of emotion is based on a combination of universal and language- and culture-specific processes. These simultaneous contributions of language and culture independent and language and culture dependent processes appear to be much stronger for vocal than for facial expression of emotion [1]. Importantly, many of the expressions used in these studies were constructed from a mono-cultural perspective, resulting in stimuli whose phonetic content is a fit for only one of the languages in the study.

In this study we focus on an as yet mostly neglected element of emotional speech: the duration of the phonetic segments in the utterance. The total duration of an emotional utterance, together with its counterpart speech rate, has traditionally received much attention in the study of the vocal expression of emotion (see for example [11, 12]. However, a more fine grained analysis of the effect of emotion on the realization of specific phonemes is a much needed contribution to the field. In addition, while the above review of the literature shows the attention of the field to cross-linguistic differences in emotion expression, the interaction of language and emotion with respect to phonetic realization has until now been ignored.

To investigate the relative contributions of universal and language specific effects on the phonetic realization of vocal emotion, we recorded and analyzed a corpus where the languages of decoders (and encoders) are fully balanced: The Demo / Kemo corpus [13]. This corpus is balanced in a number of relevant ways. First, it contains expressions by Dutch and Korean speakers and judgments by Dutch and Korean listeners. Second, in the recording phase, the same elicitation technique was used by the Korean and Dutch stage directors. Third, the speakers of both languages used the same verbal expression that was carefully constructed to contain phonemes present in both languages in combinations permissible in both languages. Finally, the emotions in the corpus are balanced in terms of valence, arousal, and dominance characteristics.

In this paper, we describe construction and validation of the Demo / Kemo corpus and present duration analyses at the phoneme level of the five most basic emotions in the corpus: Anger, Irritation (cold anger), Fear, Sadness, and Joy. We decided to analyze these five emotions as a strong test of the hypothesis that the expression of emotion is affected by language. If the expression of these basic (and often considered universal) emotions is influence by the language of the speaker, then their universality comes into question more than when, for example, the expression of pride depends on language.

2. Method

The Method section is split in three section: first, the recording of the corpus is described, second, the judgment studies that lead to the final selection of portrayals are presented and third, the phonetic segmentation and duration measurement is briefly described.

2.1. Corpus recording

In the recording of the corpus we adhered to the methods developed by Scherer and colleagues [11, 14]. This approach uses posed emotional expressions by (semi-) professional actors. While acted portrayals are in principle not spo, the approach aims at ensuring the naturalness of these expressions by using the method acting principles put forward by (see [13]). In Stanislavski's approach a director coaches the actors to produce full-blown emotional reactions by remembering and reliving a personal episode in which the target emotion occurred or by very vividly imagining such episodes. In addition, in this study, the actors were given three possible scenarios illustrating the emotion for them.

Eight Dutch actors (four males and four females) and eight Korean actors (four males and four females) participated in exchange for a small payment. All were or had been engaged in a full-time professional drama school at college level in their own country. Both directors were professionals well acquainted with the Stanislawski technique. Table 1 lists the emotions that were posed by the actors. The order in which the emotions were enacted was counterbalanced between actors. For this study, we concentrated on the following emotions: Anger, Fear, Irritation, Sadness, and Joy.

The actors had to express the emotion using a fixed phrase [nuto hɔm sɛpikaŋ]. This phrase was constructed according to the following three criteria. First, the phrase contains only phonemes that occur in both Dutch and Korean, in phonotactic combinations that are legal in both languages. Second, it is meaningless in both languages. Third, the phrase does not contain any clearly embedded words.

		Valence	
		Positive	Negative
Arousal	High	Joy Pride	Anger Fear
	Low	Tenderness Relief	Sadness Irritation

Table 1: The emotions in the corpus in a valence by arousal grid.

2.2. Judgment studies

Two judgment studies were conducted to investigate the quality and naturalness of the emotional expressions as judged by listeners sharing the native language of the actors.

2.2.1. Participants

Two groups of listeners participated in the experiment: 24 Dutch listeners (11 males, 13 females) recruited from the Radboud University Nijmegen in the Netherlands, and 24 Korean listeners (12 males, 12 females) recruited from Korea University in Seoul, Korea. All participants were students and participated in exchange for a small payment or course credits. All were native speakers of Dutch and Korean respectively and none of them reported any hearing or speech problems.

2.2.2. Materials

As described above, the materials were the 256 Dutch and 256 Korean selected utterances (8 actors * 8 emotions * 4 repetitions). They were segmented into separate wave files (mono, 44.1 kHz, 16 bit, uncompressed) that were not normalized with respect to intensity.

2.2.3. Procedure

The participants classified each of the 256 stimuli from their native language, that were presented to them in pseudo-random order. Stimuli were classified as one of the eight emotions ("anger", "fear", "sadness", "irritation", "joy", "pride", "tenderness", "relief"), or as "neutral". All response options were shown in written form on a computer screen, each in a separate square (all equally sized), at the same position (that reflected the valance and arousal properties of the stimulus) as shown in Table 1 and with the response option "neutral" in the middle. Participants indicated their response with a mouse click on the square that contained the name of the emotion category. After each categorical rating, participants had to indicate the naturalness of the expression on a scale ranging from 1 (very unnatural) to 4 (very natural).

2.2.4. Portrayal selection

We computed unbiased hit rates for each portrayal [15]. For the final corpus, the two portrayals of each actor-emotion pair with the highest unbiased hit rate were selected. When there was a tie, the portrayal with the higher naturalness rating was selected. When there still was a tie, portrayals that were confused with portrayals of the same emotion family were favored.

2.3. Segmentation and duration measurement

The portrayals of the five emotions Anger, Fear, Irritation, Sadness, and Joy were manually segmented at the phoneme level with the Praat speech analysis program [16]. All 40 portrayals (five emotions times four actors times 2 languages) were segmented by the same independent labeler. Based on these segmentations, we extracted the total duration and the duration for each phonetic segment in [nuto hɔm sɛpikaŋ]. In the Results section, the relationship between language, emotion and the duration of the phonetic segments will be presented.

3. Results and Discussion

As an initial test of the effect of language and emotion on the duration of phonetic segments, we performed a mixed analysis of variance with segment duration as dependent variable. Language (Dutch versus Korean) was entered as independent between subject variable and emotion (Anger, Fear, Irritation, Sadness, and Joy) and Segment (14 levels) were entered as independent within subject variables. This analysis revealed a main effect of language (F [1,28] = 1.076, $p < 0.001$), indicating that the Korean expressions are shorter on average. In addition, a main effect of phonetic segment was found (F [13, 364] = 132,08, $p < 0.001$) indicating that segments differ in their duration, irrespective of language. Importantly, while we did not find an overall main effect of Emotion (F [4,112] = 1.89, *n.s.*), we did find two significant interactions, one between emotion and phonetic segment (F [52, 1456] = 2.38, $p < 0.001$), and one between language and phonetic segment (F [13, 364] = 21,56 $p < 0.001$). In addition, the three-way interaction between language, emotion, and phonetic segment was also significant (F [52, 1456] = 7.13, $p < 0.001$). These three interactions illustrate the important role of both language and emotion *and* their interactions in the realization of phonetic segments.

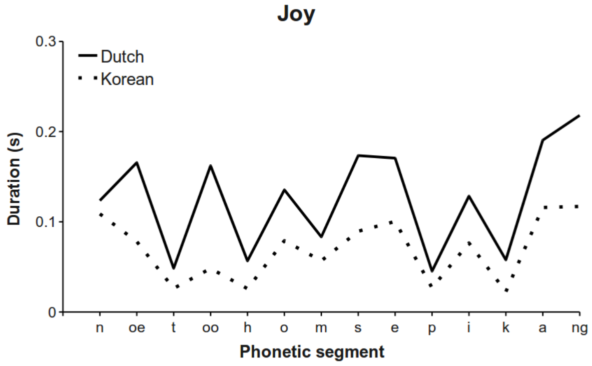

Figure 1: The duration of each phonetic segment in the carrier phrase for the emotion Joy for Dutch and Korean.

To further investigate the effect of emotion expression on segment duration, we analyzed each the effect of language and emotion on each segment separately. The charts in Figure 1 to 5 depict the duration of the segments in the carrier phrase for each emotion for Dutch and Korean. For all emotions the line representing the Dutch speakers are higher than the lines representing the Korean speakers, demonstrating the significant effect of language. The Dutch expression are not only longer, their pattern is also clearly different from that for Korean for all emotions.

Together, the figures are indicative of the effect of language and emotion on segment duration. To statistically investigate these effects, we conducted separate analyses of variance for each phonetic segment with language and emotion as independent variables and segment duration as dependent variable. The main effect of language was significant for all segments (F_{min} = [1,30] = 11.86, $p < 0.002$) with the exception of the first /n/ and the /a/ at the end. Table 2 depicts the F values, degrees of freedom and significance for the effect of emotion and the interaction between emotion and language on the duration of each segment.

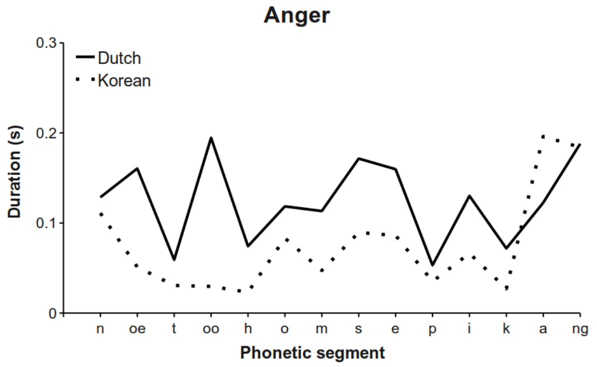

Figure 2: The duration of each phonetic segment in the carrier phrase for the emotion Anger separately for Dutch and Korean.

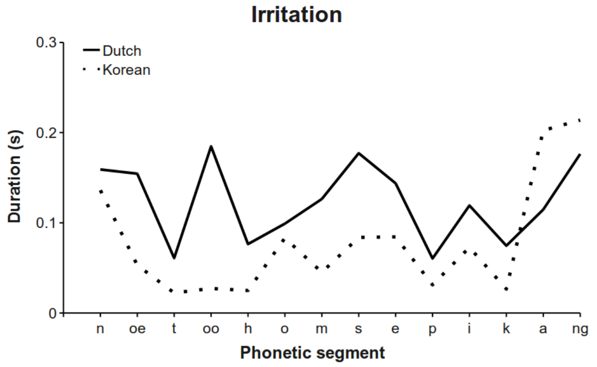

Figure 3: The duration of each phonetic segment in the carrier phrase for the emotion Irritation for Dutch and Korean.

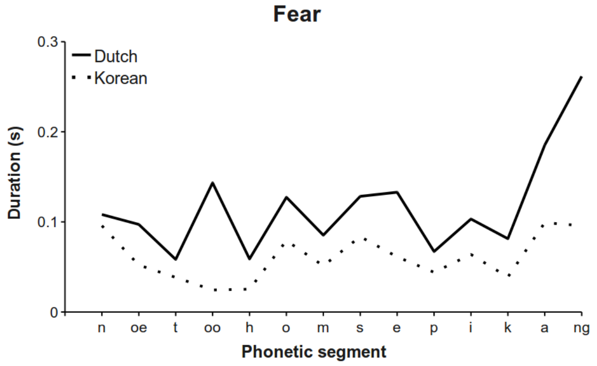

Figure 4: The duration of each phonetic segment in the carrier phrase for the emotion Fear for Dutch and Korean.

The statistics in Table 2 show that emotion affects the phonetic realizations of speech and that this influence is often language specific. The analyses show that for four segments (/t/, /o/, /h/, and /O/) in the second and third syllable of the nonsense sentence, emotion did not affect duration. For the segments /n/, /p/, and /k/, emotion affected duration for Korean and Dutch in

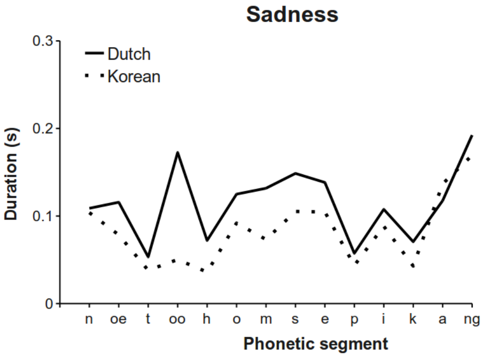

Figure 5: The duration of each phonetic segment in the carrier phrase for the emotion Sadness for Dutch and Korean.

	Emotion			Emotion x Language		
Segment	F	df	p	F	df	p
n	4.38	4,120	0.002	0.15	4,120	*n.s.*
u	9,98	4,120	0.000	9.23	4,120	0.001
t	1.50	4,120	*n.s.*	1.69	4,120	*n.s*
o	1.56	4,120	*n.s.*	2.06	4,120	*n.s*
h	1.95	4,120	*n.s.*	1.60	4,120	*n.s.*
ɔ	1.15	4,120	*n.s.*	2.23	4,120	*n.s.*
m	6.67	4,120	0.000	4.49	4,120	0.002
s	3.17	4,120	0.016	3.79	4,120	0.001
ɛ	9.65	4,120	0.000	3.35	4,120	0.01
p	3.43	4,120	0.011	0.55	4,120	*n.s.*
i	4.03	4,120	0.006	5.01	4,120	0.001
k	4.64	4,120	0.002	1.32	4,120	*n.s.*
a	1.22	4,120	*n.s.*	10.93	4,120	0.000
ŋ	0.58	4,120	*n.s.*	12.10	4,120	0.001

Table 2: Statistics indicating the significance of the main effect for emotion and for the interaction between language and emotion on the duration of each individual segment. The Bonferroni level for multiple testing lies at $\alpha < 0.004$.

similar ways. For the remaining seven segments, however emotion affected duration in Korean and Dutch differentially. In the present study, the influence of emotion and language, and their combined influence occurs mostly at the middle and the end of the utterance, and mostly for sonorous segments. Whether these effects hold up in other languages and other emotions, is subject for further study.

4. Conclusion

This paper presented an exploratory study into the relationship among language, emotion, and phonetic realization. To this end, we recorded emotional speech from Dutch and Korean actors and selected their best portrayals by means of an emotion judgment study. We then measured the duration of each individual phonetic segment and showed that language as well as emotion has a significant influence on the duration of a segment. Importantly, these two factors often interact with one another, thus creating language specific effect of emotion on phonetic segment duration.

5. Acknowledgements

Martijn Goudbeek was supported by funding from the VICI project “Bridging the gap between psycholinguistics and computational linguistics: the case of referring expressions”, funded by the Netherlands Organisation for Scientific Research (NWO grant 277-70-007, awarded to Emiel Krahmer). Mirjam Broersma was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO). We thank Jiyoun Choi for her help recording the Korean corpus and testing the Korean participants, and Sammie Tarenskeen for testing the Dutch participants and her help in labeling the phonetic segments.

6. References

[1] Elfenbein, H.A. and Ambady, N. “On the universality and cultural specificity of emotion recognition: A meta-analysis”, Psych Bull., 128:203-235, 2002.

[2] Ekman, P., Sorenson, R.E., and Friesen, W.V. “Pan-Cultural Elements in Facial Displays of Emotion”, Science, 164:86-88, 1969.

[3] Ekman, P, “An Argument for Basic Emotions.”, Cogn Emotion., 6:169-200, 1992.

[4] Albas, D.C., McCluskey, K.W., and Albas, C.A.“Perception of the emotional content of speech: a comparison of two Canadian groups”, J, Cross Cult Psychol., 7:481-489, 1976.

[5] van Bezooijen, R., Otto, S.A., and Heenan, T,“Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics”, J Cross Cult Psychol., 14:387-406, 1983.

[6] Scherer, K.R., Banse, R., Wallbott, H.G., “Emotion inferences from vocal expression correlate across languages and cultures”, J Cross Cult Psychol., 32: 76-92, 2001.

[7] Thompson, W.F. and Balkwill, L-L., “Decoding speech prosody in five languages”, Semiotica, 158:407-424, 2006.

[8] Pell, M.D. and Skorup, V, “Implicit processing of emotional prosody in a foreign versus native language”, Speech Comm., 50:519-530, 2008.

[9] Sauter, D., Eisner, F., Ekman, P., and Scott, S.K., “Universal vocal signals of emotion”, Proc of the 31st Annual Meeting of the Cognitive Science Society. Amsterdam, The Netherlands, 2009.

[10] Pell, M.D., Monetta, L., Paulmann, S., and Kotz, S., “Recognizing Emotions in a Foreign Language”, J Nonverbal Behav., 33:107-120, 2009.

[11] Banse, R., and Scherer, K.R. “Acoustic profiles in vocal emotion expression”, J Pers Soc Psychol., 70:614-636, 1996.

[12] Juslin, P., and Laukka, P. “Communication of emotions in vocal expression and music performance: Different channels, same code?”, Psych Bull., 129:770-814, 2003.

[13] Goudbeek, M. and Broersma, M. “The Demo / Kemo corpus: A principled approach to the study of cross-cultural differences in the vocal expression and perception of emotion”, Proc of LREC 2010, valetta, Malta, 2010.

[14] Bänziger, T. and Scherer, K.R., “Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus”, ACII, 476-487, 2007.

[15] Wagner, H.L., “On measuring performance in category judgment studies of nonverbal behavior”, J Nonverbal Behav., 17:3-28, 1993.

[16] Boersma, P., “Praat, a system for doing phonetics by computer.”, Glot International, 5:341-345, 2001.