BA Thesis

Emotion Perception in Adverse Listening Conditions

Emotion Perception in Italian Speech in Noise by Dutch Listeners

Jiska Koemans

4366360

25-08-2016

Odette Scharenborg (supervisor)

Contents:

Abstract

Much (cross-linguistic) research has been done on nonverbal emotion perception (e.g., the perception of laughter and facial expressions). Another important aspect of emotion perception is verbal emotion perception. Nonverbal emotion perception has been found to contain specific universal characteristics which make it possible to recognise emotions across different cultures. Research shows that verbal emotion perception may also contain universal characteristics. If so, people should be able to determine which emotion is being expressed in an utterance regardless of whether they speak the language or not. Emotion perception is typically investigated in optimal laboratory conditions. However, in everyday listening conditions, background noise is prevalent. Moreover, in today's multilingual society, many people regularly communicate in a language other than their mother tongue. This study therefore investigates verbal emotion perception in less than optimal conditions. Specifically, this study investigates the effect of not knowing the language and the effect of the presence of background noise on emotion perception.

To investigate these questions, Dutch listeners who have no experience with Italian have participated in an emotion recognition task. In the future, Italian listeners will be tested as well. The experimental stimulus set consisted of 120 emotionally-coloured utterances selected from the Italian acted EMOVO Corpus. This selection consisted of 10 different utterances (five nonsense and five semantically neutral utterances) produced in one of five emotions, i.e., anger, sadness, joy, fear, or a neutral state, by both female and male speakers. Utterances were presented in the clean, and crucially, in two babble noise conditions at 5dB and -3dB. Each utterance-emotion combination was presented only once to each listener, and thus each utterance-emotion combination only appeared in one listening condition. Utterances, emotions, and listening conditions were randomized and counterbalanced over participants. After listening to each utterance, listeners had to determine which of the five emotions they thought the utterance conveyed. The emotion decisions were compared between the two listener groups and between the different listening conditions.

In line with previous cross-linguistic emotion perception research, I found that Dutch participants were able to recognise emotions in Italian. Furthermore, a negative effect of noise on the perception of emotions was found. The perception of emotions in Italian became more difficult for Dutch listeners when listening in the -3dB SNR condition. Crucially, this only applies for anger and fear. Furthermore, effects of sentence type, speaker and trial were found for specific emotions, indicating that these factors may influence the process of emotion perception as a whole. The findings indicate that noise does influence the perception of emotions in a foreign language. However, to what extent noise influences the perception and if this really is only applicable to specific emotions should be further investigated. A follow-up study will focus on emotion perception by native Italians, and compare these results with the results of the current study.

1. Introduction

Emotion perception is one of many important aspects of human communication. In the process of communication, emotions help convey and understand messages. If emotions are interpreted incorrectly, the complete message could come across the wrong way (leading to arguments, for example). Emotions can be conveyed through facial expressions and other nonverbal features (e.g., gestures). Moreover, emotions can be perceived through speech. Emotion perception through speech has been receiving increasing attention in the linguistic research field in the past years. Nonverbal emotion is said to contain specific universal characteristics which make it possible to recognise emotions across different cultures (e.g., Ekman, 1992). Research shows that verbal emotion may also contain universal characteristics (e.g., Scherer et al., 2001). This means that people should be able to determine which emotion is expressed in an utterance regardless of whether they speak the language or not, even without any nonverbal cues. Speech and emotion perception is typically investigated in optimal laboratory conditions. However, in everyday listening conditions, background noise is prevalent. Background noise is an important influence on the perception of speech, since it negatively influences the acoustic signal (e.g., Lecumberri et al., 2010), which may cause the perception of emotions in speech to be negatively influenced as well. Moreover, in today's multilingual society, many people regularly communicate in a language other than their mother tongue. Therefore people need to be able to interpret emotions in a non-native language. For these reasons, this study investigates verbal emotion perception in less than optimal conditions. Specifically, this study investigates the effect of not knowing the language and the effect of the presence of background noise on emotion perception.

With his evolutionary theory of emotions, Darwin already described universality as a characteristic of emotions (Darwin, 1872). The psychologist Paul Ekman, among others, discovered that characteristics of certain emotions and facial expressions are identical across cultures (e.g. Ekman, 1971, 1992, 1999; Izard, 1994). Through cross-linguistic research, he identified six emotions as being universal: anger, fear, sadness, disgust, joy and surprise (Ekman, 1992). These emotions were all recognised as being the same emotions (meaning anger was recognised as anger) by members from different cultures. Ekman proposed nine characteristics an emotion should contain for the emotion to be classified as universal. These nine characteristics are used to identify nonverbal emotions (see Ekman, 1992, for a detailed description of each universal sign).

The extent to which verbal emotion perception happens through universal or cultural principles is not completely clear. Research has found evidence for the existence of universal cues in verbal emotions (e.g. Scherer et al., 2001; Thompson & Balkwill, 2006; Pell et al.,

2009; Da Silva et al., 2016). These studies investigated verbal emotion perception by non-native listeners (who did not have any knowledge of the foreign language). All studies showed above chance correct emotion recognition rates for all languages, meaning that non-native listeners are able to recognise emotions in a language they do not speak. This indicates that verbal emotion expression contains universal features, which makes it possible for members from different cultures, speaking different languages, to identify emotions cross-linguistically. Nevertheless, comparing the verbal emotion perception accuracies of non-native listeners with those of native listeners showed significantly better recognition rates for verbal emotion perception in native listening (Scherer et al., 2001; Thompson & Balkwill, 2006 and Pell et al., 2009). Identifying emotions in a native language may have certain advantages, based on linguistic and cultural aspects, because natives are able to use these language and cultural specific aspects to identify emotions, on top of universal cues (Pell et al., 2009). Note, however, that the study of Da Silva et al. (2016) partially focussed on the influence of native language, culture and emotional experiences on the perception of emotions (in addition to cross-linguistic verbal emotion perception) and their results showed no significant influence of these cultural aspects on verbal emotion perception. According to Scherer and colleagues (2011), despite a large amount of empirical research on the topic of cross-cultural verbal emotion perception and its universality, no hard conclusions can be drawn yet. However, based on existing literature, they do think compelling evidence is found that points towards the existence of universality in decoding emotions cross-culturally (Scherer et al., 2011). Still, more evidence is needed to be able to answer the questions regarding universality in verbal emotion perception.

In natural circumstances, most conversations occur in areas where some kind of background noise is present (often the sound of other people talking). Therefore it may be important to take this natural factor into account while investigating verbal emotion perception. So far, no studies have focused on verbal emotion perception in noise by human listeners. However, noise is expected to negatively influence the perception of emotions in speech, because noise negatively affects the acoustic speech signal in general (Lecumberri et al., 2010). Furthermore, noise affects non-native listening more severely than native listening, possibly indicating that non-native emotion perception will be influenced by noise as well (more than emotion perception by natives). However, background noise seems to affect the perception of prosody speech less severely than speech itself, as observed by Van Zyl and Hanekom (2011). They found that the perception of prosody did not significantly deteriorate when listening in noise, in contrast to the perception of speech. Since the prosody of emotions differs from the prosody of normal speech, the way background noise affects the perception of

emotional speech could differ from the way it affects normal speech and one could assume emotional speech would still be recognised in noisy environments. However, given the fact that previous studies focused on prosody in speech in general (in noisy environments), rather than the prosody of emotions and emotional speech, it is not clear yet in what way and in how far verbal emotion perception will be influenced by noise.

A few previous studies have focused on verbal emotion perception in noise (e.g., Schuller et al. 2006; Zhao et al., 2013), but these studies used computers and statistical models (e.g., support vector machines (SVM) and artificial neural networks (ANN)) rather than human judges to identify emotions in a noisy environment. The experiments focused on testing in how far different classification models (e.g., sparce representation classifier (SRC)) are able to be trained to identify and classify emotions in speech. Moreover, these studies used white noise as background noise. White noise is not a noise typically encountered in natural conversational circumstances. Therefore, babble noise was used in the current study, which has more similarities with natural background noise in conversational environments compared to the noise used in earlier studies, which allows me to study verbal emotion perception in natural circumstances. The studies by Schuller and colleagues and Zhao and colleagues did find a negative influence of noise on the perception of emotions in speech, meaning that recognition is more difficult when noise is present. Despite the difference in their approaches and the approach of the current study, this might be an indication that background noise will negatively affect verbal emotion perception by human listeners as well.

For the reasons mentioned above, the current study focuses on non-native verbal emotion perception in noise by human listeners, using babble noise. The experiment was carried out to answer two questions: in how far are Dutch people able to perceive emotions in a language they do not speak; and what is the effect of noise on non-native verbal emotion perception? I expect to find a negative effect of noise on the perception of the emotions in the different listening conditions, meaning that the emotions should be recognised worse when listening in noise. Moreover, I expect to find a differential effect of noise on the perception of the different emotions, based on the fact that the prosody and acoustic features of the investigated emotions differ (e.g., Sobin & Alpert, 1999).

To answer the experimental questions, an experiment was carried out in which Dutch participants judged Italian emotional speech in a clean condition as well as in two noise conditions. Ideally, I wanted to investigate native verbal emotion perception in noise as well, by testing Italians in the same experimental task. However, for the current study it was not possible to find enough Italian participants within the timespan of the project. For this reason

I plan to continue this experiment in the future with Italian participants, and I will compare the results of the future study with those of the current study.

For the experiment, Italian emotional speech was extracted from the EMOVO Corpus (Constantini et al., 2014). The EMOVO Corpus was chosen because Italian is a language most Dutch people do not speak. Moreover, the corpus is available for downloading online, after which the stimuli could be manipulated for the noise conditions. The emotional speech in the corpus is acted. Using actors to portray emotional speech may lead to utterances being overacted, meaning that prototypical cues are used (and may be exaggerated) to portray a specific emotion, which may make utterances by actors less useful (Wilting et al., 2006). However, for the current study, I will still be able to draw conclusions on the effect of noise on emotion perception by using acted speech. Therefore, I will consider the fact that the speech is acted, but I can use the speech nevertheless.

2. Methods

2.1 Participants

26 Dutch speakers, who did not speak or understand Italian, participated in the experiment. All were students at the Radboud University, Nijmegen, The Netherlands. Their mean age was 21.8 years (*SD* = 3.5; 9 males; 17 female).

2.2 Materials

2.2.1 EMOVO Corpus

The stimuli used for the experiment were taken from the EMOVO corpus (Constantini et al., 2014). The EMOVO corpus is an Italian corpus, which consists of 588 recorded Italian emotional utterances portraying six emotions and a neutral state (Constantini et al., 2014). These emotions are anger, fear, sadness, joy, surprise and disgust. Six professional actors (age range 23-30 years), three male and three female, recorded these emotions in 14 emotionally neutral sentences. Nine of these were semantically neutral (e.g., ‘workers get up early’) and five were ‘nonsense’ sentences (with correct grammar, e.g., ‘the strong house wants with bread’). The actors were selected based on their acting background, specifically on their knowledge of the ‘Stanislavsky method’ (Constantini et al., 2014). Through this method actors evoke emotions by recalling real life situations where a specific emotion was experienced.

The corpus is available for downloading online through the following link: http://voice.fub.it/activities/corpora/emovo/index.html

The corpus was recorded in Italy, at the laboratories of the Fondazione Ugo Bordoni and validated through a superficial validation test. The focus of the validation was on the question whether the actors were able to portray the emotions, not to study in how far Italians are able to recognise emotions in their native language. Twelve native Italians participated in the validation test (Constantini et al., 2014). The test consisted of a listening task in which the participants had to choose from two options which emotion they heard. They had to listen to 84 recorded utterances: for each of the six different actors who were recorded for the corpus, two nonsense utterances per actor were used, in each of the seven emotional states. Based on this validation test, it was concluded that the actors were all able to portray the emotions, because the emotional portrayals of all actors were recognised at above chance rates.

From the original corpus, a subset of 100 recorded utterances was chosen to be used in the current experiment. These utterances consisted of ten different sentences, each portrayed in five emotions: anger, sadness, fear, joy and neutral. Moreover, every sentence-emotion pair was represented twice, once by a female and a male speaker, yielding 100 recorded utterances in total.

2.2.2 Speaker selection

Studies on the perception of the emotions in this corpus showed great variety in the actors' ability to portray the emotions (Constantini et al., 2014, Giovannella et al., 2009; 2012). Following the findings of the studies by Giovannella and colleagues, the female and male actors who were found to portray a specific emotion best were chosen to represent that emotion in the subset used in this study. Table 1 shows which speakers were chosen to represent each specific emotion, and the recognition rate obtained by Giovannella and colleagues (2009, 2012). F or M represents the gender of the speaker (respectively female and male) and the number represents if it was the first, second or third speaker (since three actors have been recorded per gender).

Table 1: Speaker and recognition rate per speaker per emotion

Emotion	Speaker	Recognition rate (%) Female	Recognition rate (%) Male
Anger	F2, M2	53	82
Fear	F2, M1	100	65
Sadness	F1, M1	70	44
Joy	F2, M2	71	23
Neutral	F3, M3	37	29

Table 2: The Italian utterances used in the experiment with their English translation

Italian	English
Gli operai si alzano presto.	Workers get up early.
I vigili sono muniti di pistola.	Firefighters are equipped with a gun.
La cascata fa molto rumore.	The waterfall makes a lot of noise.
Ora prendo la felpa di là ed esco per fare una passeggiata.	Now I take the sweatshirt and go for a walk.
Un attimo dopo s'è incamminato ... ed è inciampato.	A moment later he had walked ... and stumbled.
La casa forte vuole col pane.	The strong house wants with bread.
La forza trova il passo e l'aglio rosso.	The force is up and red garlic.
Il gatto sta scorrendo nella pera.	The cat is flowing in pear.
Insalata pastasciutta coscia d'agnello limoncello.	Pasta salad leg of lamb limoncello.
Uno quarantatré dieci mille cinquantasette venti.	One forty-three ten thousand fifty-seven twenty.

2.2.3 Sentence selection

In the original corpus, 14 different sentences were used to capture each of the emotion, five of which were nonsense utterances and nine were semantically normal. The sentences were chosen such that their meaning did not interfere with the emotion that was portrayed. The sentences were therefore emotionally neutral in content.

For the current experiment, of the nine semantically normal sentences in the corpus, four were excluded based on their form and/or content. Two sentences were questions and two contained names. Since the prosody of questions is different from the prosody of statements, it was decided to leave out the questions from the subset, to avoid a possible influence of prosody. Moreover, names may evoke certain memories or feelings, for example because participants may know someone with that name or it makes an utterance more personal in general. Therefore, the sentences containing names were left out as well. The

sentences used in this experiment, thus consisted of five semantically normal and five nonsense sentences. Because in first instance the experiment was designed to test both Dutch and Italian participants, I decided to use both nonsense and normal utterances to be able to determine any interference from content for the Italians. Table 2 shows the Italian utterances that were used from the original corpus and their English translation (Constantini et al., 2014).

2.2.4 Listening conditions

All participants were tested in three conditions: a clean and two babble noise conditions. Not all items were equally loud, and differences in loudness might influence the outcome of the experiment. Therefore, the loudness of all items needed to be equalised. To do this, I used *peak normalisation* in *Praat*, with a *Praat* script (Kerkhoff, 2015). *Peak normalisation* changes the maximum amplitude (or maximum peak) of a speech signal, to make it louder or less loud. In this case, I needed (most of) the items to be louder, so I needed to enlarge the peaks. I enlarged the maximum amplitude of all items to 97 dB (standard setting in the script), which is quite loud, not only for all items to be equally as loud, but also to make sure the items were loud enough for all participants.

Subsequently, 8-speaker babble noise was added with *Praat* to the original files at two SNRs. Babble noise was chosen, because this type of noise is common in everyday communication. The babble noise used in this experiment consisted of Dutch speech from four different native Dutch speakers (Scharenborg et al., 2014). The speakers produced isolated words. To create the 8-speaker babble, two random words from each of the four speakers were mixed together.

The SNRs were determined using a pretest. For the pretest, babble noise at an SNR of 5dB and -3dB were tested. Two subjects performed the emotion perception experiment in the clean condition as well as the two noise conditions. The participants had to carry out the same task as used in the main experiment (see Section 2.3.1). After the experiment, they were asked if they noticed anything specific about the items and if there was anything distracting or unpleasant about the items. They stated that some of the items, in particular the sadness items of the male speaker, contained background noise on top of the babble noise, which was prominent despite the presence of the babble noise. The noise was present throughout the recordings. Removing the noise would result in removing part of the speech signal as well, which would result in speech that would sound less natural. For this reason, the noise was suppressed instead of removed. The noise was suppressed by using *noise suppression* in *Audacity* (Audacity Team, 2016). By analysing a part of the signal that contains noise, the program is able to suppress the noise in the complete speech signal without adjusting the

speech itself. As a result, the noise becomes less noticeable while the speech still sounds natural. After suppression of the noise, a second pretest was carried out. The same participants were tested as in the first pretest to make sure they were able to compare the first and second test. They had to do the exact same task and afterwards they were asked again whether they noticed anything specific about the items. This time, they stated that they were no longer aware of the noise. The items used in the second pretest were subsequently used in the main experiment.

Analysis of the emotion recognition accuracies of the first pretest showed that at 5 dB the participants could still rather easily distinguish the emotion (and the utterance) while in the -3 dB condition this was (as expected) more difficult but not impossible. The SNRs of -3 and 5 were deemed suitable for the experiment. The listening conditions used in the experiment were thus: clean, noise SNR 5 dB and noise SNR -3 dB. The condition with an SNR of 5 dB represents a listening condition in which the noise is distracting but does not interfere too much with perceiving the emotion and speech, whereas the -3 dB condition represents a listening condition in which the noise is much more dominant and recognising the emotion and speech is much harder.

2.2.5 Experimental lists

Twelve experimental lists were composed to test the stimuli as randomly as possible. Each participant was tested on one experimental list, consisting of 120 items. Each experimental list was divided into three 40-item subsets, one for each of the listening conditions. Each subset (or listening condition) contained eight utterances for each of the five emotions: four utterances from the selected female speaker and four from the selected male speaker. Per speaker, two utterances were of semantically normal content and two were nonsense utterances. Each sentence-emotion pair appeared only once per 40-item subset.

The randomisation was done by changing the order in which the listening conditions appeared, and by changing the order in which the 40-item subsets appeared. For the first six lists the order of these subsets was block 1 – block 2 – block 3. The order of the listening conditions differed. For instance, the order of the listening conditions in list 1 was clean – noise 5dB – noise -3dB, and the order of list 2 was noise 5dB – noise -3dB – clean, and so on until all six possible orders were used. Thus, to be able to randomise the remaining lists, the order of the subsets was changed: block 2 – block 3 – block 1. Therefore, the order in which the utterances appeared was different in these lists. Furthermore, the same randomisation of the listening conditions was used. Because this was a fairly small experiment, only these 12

lists were used. However, in a follow-up study it is possible to further randomise the lists by using different orders in which the subsets appear.

2.3 Procedure

2.3.1 Experimental task

Participants were tested in soundproof booths in the experimental lab in the Centre for Language Studies in the Erasmus building of the Radboud University, Nijmegen wearing headphones. Participants listened to 120 utterances and for each utterance had to determine which of five emotions they heard. These five options, anger, fear, sadness, joy and neutral, were shown on a computer screen and for each of the emotions a corresponding key on the keyboard was labelled. The participants had to press the key representing the emotion they thought they heard. Every utterance was played once and by pressing one of the marked keys the next utterance would automatically play after a short pause.

2.3.2 Instructions

The participants were told not to pay attention to the content of the utterances. They had to focus on the emotion they heard. This was to make sure the Dutch participants did not try to give meaning to the words and sentences they heard and that they would not let themselves get distracted if they might recognise something from the content.

2.4 Data analysis

To find out in how far Dutch people are able to recognise emotions in Italian speech, the recognition rates were calculated for each emotion. Because this was done by comparing how often participants were correct or incorrect in their response, the dependent variable of the analysis was whether the participant gave a correct response.

Furthermore, listening condition (clean, noise 5dB and noise -3dB) was added as a fixed factor (with clean on the intercept), to study whether noise influences the perception of emotions in speech and if so, in how far noise influences the perception. Thus, I investigated if noise can predict whether an emotion will be recognised correctly or incorrectly.

To control for possible influences of other factors, sentence type, speaker and trial were added as fixed factors as well. I studied whether the type of the sentence (normal vs. nonsense) in any way influences the perception of the emotions (with the normal sentences on the intercept). Furthermore, I investigated whether the different speakers influence the perception of the emotions (with the female speaker on the intercept), and if the perception

changes when a participant judged more trials (meaning they progress in the experimental task). Subject and stimuli were added as random factors, because of their possible individual-specific influence and variance. For the fixed factors, interactions will be looked at as well.

All data will be analysed using generalized linear mixed-effect models (e.g., Baayen et al., 2008) in R (R development core team, 2011) containing fixed factors as well as random factors. For each emotion, the best fitting model will be built based on the fixed and random factors. Starting with a model containing all predictors and possible interactions, the best model will be found using a backward stepwise selection procedure. This means that the model will be reduced by removing the interactions and predictors that are not significant (removing the least significant first), until the model is found that contains only significant predictors and interactions.

3. Results

The analyses were carried out to find out in how far Dutch people are able to recognise emotions in Italian and whether noise influences the recognition. Furthermore, the possible effects of sentence type, speaker and trial were investigated.

Table 3 shows the overall recognition rate for each emotion. T-tests showed that all emotions were recognised significantly better than chance ($p < .001$ for every emotion). For each emotion, a total of 576 utterances were analysed. Furthermore, table 3 shows the proportion of the false positives, which indicates, for instance, how often an emotion was judged as being neutral when in fact it was another emotion. The higher the false positive rate, the more often the particular emotion was wrongly chosen. The false positive rates, however, were not taken into account in the analyses; they were only calculated to help get a better idea of what the recognition rates indicate.

The overall results in table 3 show that Dutch participants were able to recognise emotions in Italian. Anger and neutral were recognised best. Fear and sadness were recognised less well and joy was recognised the worst. Anger was also the emotion with the lowest false positive rate, followed by joy, fear and sadness (respectively). Whereas neutral was recognised best, the false positive rate for neutral was the highest. The high false positive rate of neutral indicates that it is often wrongly chosen, or confused with other emotions, despite its high recognition rate. The low false positive rate of anger indicates that it is very rarely wrongly chosen, in addition to a high recognition rate.

Figure 1a shows the recognition rates per emotion per noise condition, and figure 1b shows the false positive rates per emotion per noise condition. Figure 1a shows a decrease in

the recognition rates for each emotion when listening in the different noise conditions, with a greater decrease when listening in the noise condition with the most background noise (-3dB).

Figure 1b shows an increase in the false positive rates as the SNRs increase, which too indicates that the perception of emotions becomes harder when listening in noise, because participants get confused more often when judging emotions in noise.

Figure 2 shows the recognition rates per sentence type per emotion. For anger and joy, the recognition rate of the nonsense utterances is higher than that of the normal utterances. For fear and sadness the recognition rate of the normal utterances is higher compared to that of the nonsense utterances, and for neutral the recognition rates are equal.

In figure 3 the recognition rate per speaker per emotion is shown. The recognition rate for the male speaker was higher only for anger; for fear, sadness and joy the recognition rate of the female speaker was higher. For neutral, the recognition rates of both speakers were equal.

Table 3: recognition rate and false positive rate per emotion

Emotion	Recognition rate (%)	False positive (%)
Anger	89.4	9.3
Fear	53.5	33.9
Sadness	55.9	38.2
Joy	45.7	22.0
Neutral	89.6	47.8

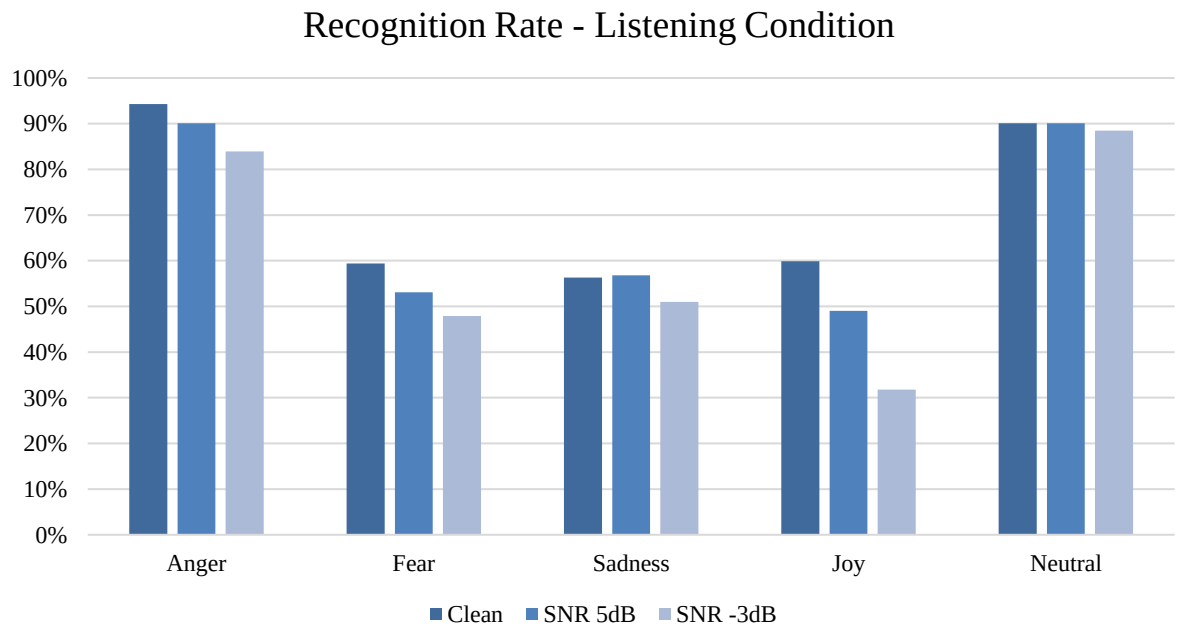


Figure 1a: Recognition rate per emotion per listening condition

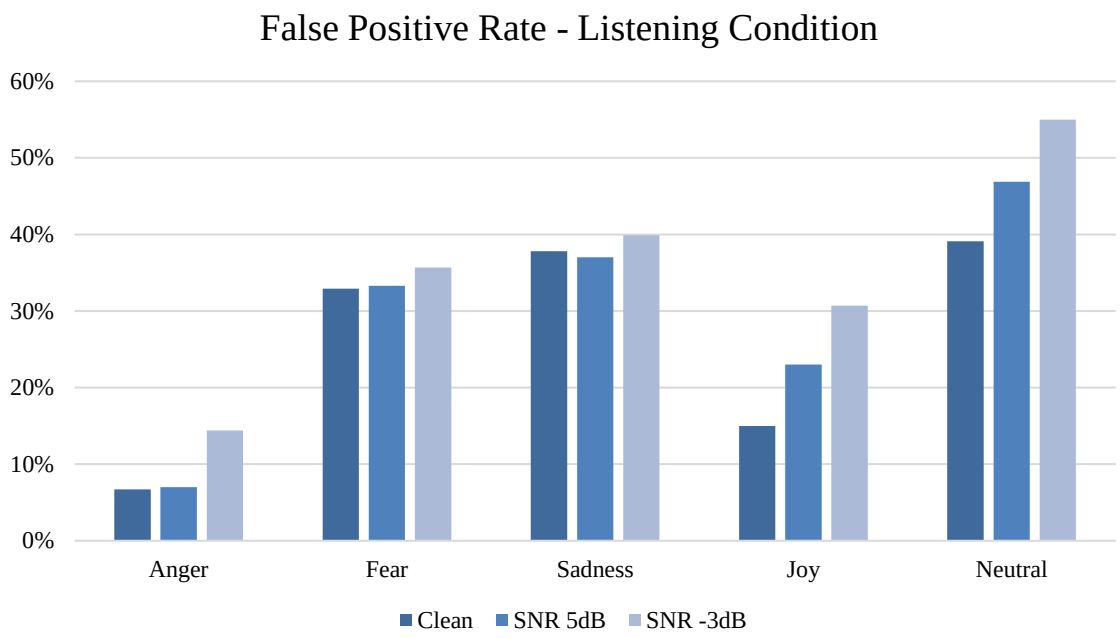


Figure 1b: False positive rate per emotion per listening condition

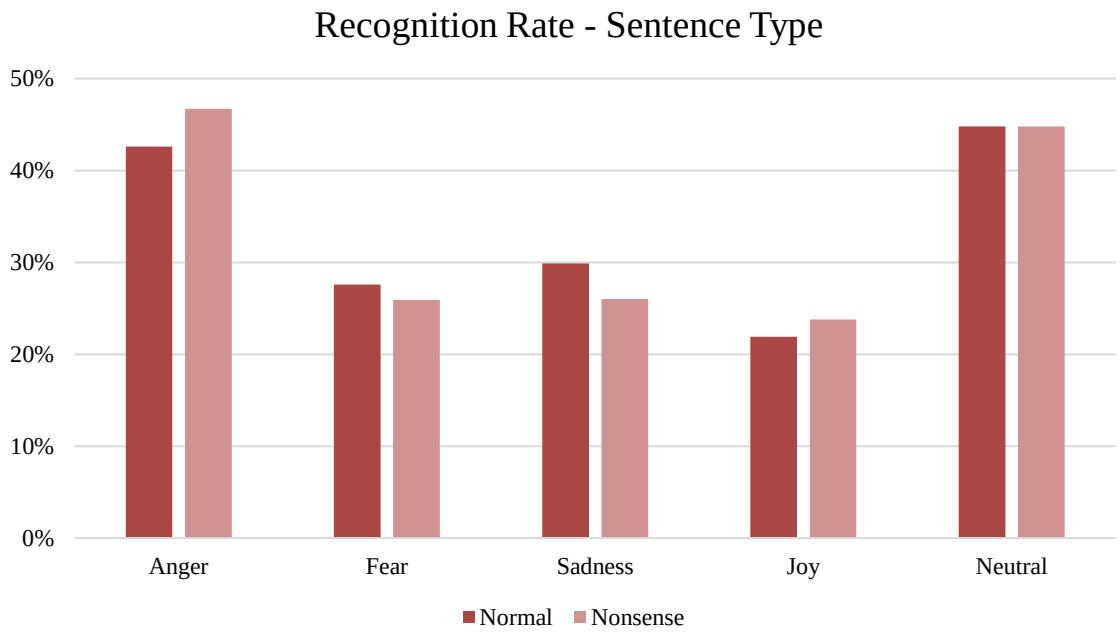


Figure 2: Recognition rate per emotion per sentence type

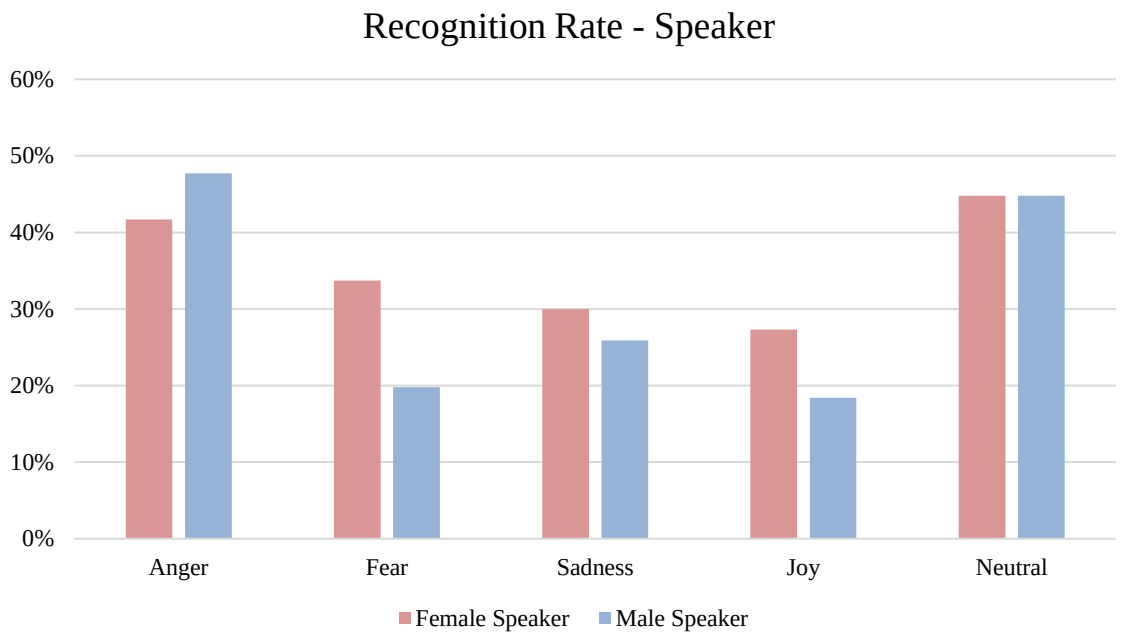


Figure 3: Recognition rate per emotion per speaker

3.1 Anger

Overall, anger was the best recognised emotion, as shown by figure 1a (see also *Recognition Rate* in table 3), especially if you take into account that it was rarely wrongly chosen (9.3%), which is shown by figure 1b (see also *False Positive* in table 3). Figure 1a shows a decrease in the recognition rate for anger when listening in both noise conditions, compared to the clean condition. Anger was less often recognised correctly when listening in SNR 5 dB (90.1%) and SNR -3 dB (83.9%), compared to the clean condition (94.3%). Only the recognition rate of the -3 dB condition is found to significantly differ from that of the clean condition (see *SNR - 3 dB*, table 4). When identifying emotions while listening in the -3 dB SNR condition, where the noise was the loudest, participants recognised anger less well compared to listening in the condition without noise. This means that more severe noise negatively influences the recognition of anger by Dutch listeners when identifying emotions in Italian.

Figure 2 shows an effect of sentence type for anger. Anger was more often correctly recognised for nonsense sentences (46.7%) than for normal sentences (42.7%). This effect was found to be significant (see table 4), meaning that participants recognised anger better when identifying emotions in the nonsense utterances compared to the normal utterances.

Last, figure 3 shows an effect of speaker for anger. The observed difference between the male and female speaker was found to be significant. When listening to the male speaker (47.7%), participants more often recognised anger correctly compared to listening to the female speaker (41.7%).

Table 4: Results for anger

Fixed factor	Estimate (β)	Std. Error	*p*
(Intercept)	1.29	.65	.048
Nonsense Sentence	.94	.94	<.01
Male Speaker	1.55	.37	<.001
SNR 5 dB	-.67	.45	.14
SNR -3 dB	-1.31	.43	<.01

3.2 Fear

Overall, fear was correctly recognised a little over 50 percent of the time, with a false positive rate of 33.9% (see table 3). Furthermore, figure 1a shows a decrease in the recognition rates for SNR 5 dB (53.1%) and SNR -3 dB (47.9%), compared to the clean condition (59.4%). However, only the observed difference between SNR -3 dB and clean was found to be significant (*SNR -3 dB*, table 5), meaning that relatively severe noise negatively influences the recognition of fear by Dutch listeners when identifying emotions in Italian.

Furthermore, as shown by figure 3, a difference was found in the recognition rates for the utterances by the male speaker compared to the female speaker. The observed difference was found to be significant. The recognition rate for the male speaker portraying fear (19.8%) was significantly lower than that of the female speaker (33.7%), which means that the male speaker was worse in portraying fear than the female speaker (see also table 5, *Male Speaker*).

Table 5: Results for fear

Variable	Estimate (β)	Std. Error	P-value
(Intercept)	2.02	.69	<.01
Male Speaker	-1.27	.23	<.001
SNR 5 dB	-.27	.27	.32
SNR -3 dB	-.55	.27	.046

3.3 Sadness

Overall, sadness was correctly recognised a little over 50 percent of the time, with a false positive rate of 38.2% (see table 3). In figure 1a, a very small increase is visible in the recognition rate for sadness when listening in SNR 5 (56.8%), and for listening in SNR -3 (51%) a slight decrease is visible, both compared to the clean condition (56.3%). Furthermore, a slight decrease in the recognition rate is found for sadness when identifying emotions in nonsense sentences (26%) compared to normal sentences (29.9%) (figure 2), and a slight

decrease is shown in the recognition rate when listening to the male speaker (25.9%) compared to the female speaker (30%) (figure 3), but none of these overall effects are found to be significant.

For trial, however, a significant effect has been found. Participants correctly recognised sadness utterances significantly more often as they progressed in the experimental task, which means that the perception of sadness becomes easier as you hear it more often.

Table 6: results for sadness

Emotion	Variable	Estimate (β)	Std. Error	P-value
Sadness	(Intercept)	-1.06	.50	.03
Sadness	Trial	.01	.003	<.01

3.4 Joy

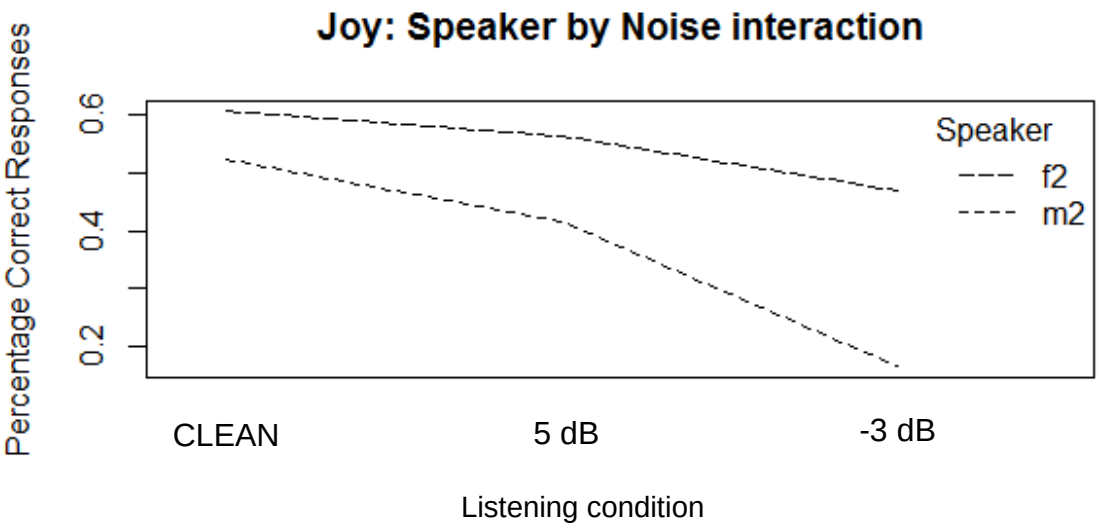
Overall, joy was recognised less well compared to the other emotions, with a recognition rate of 45.7% (table 3). However, at 22%, the false positive rate for joy is quite low compared to that of the other emotions (except anger). Figure 1a shows a decrease in the recognition rate for joy when listening in SNR 5dB (49%) and SNR -3 dB (31.8%), compared to the clean condition (59.9%). Furthermore, figure 2 shows a slight increase in the recognition rate when identifying emotions in the nonsense utterances (23.8%) compared to the normal utterances (21.9%), and figure 3 shows a decrease in the recognition rate when listening to the male speaker (18.4%) compared to the female speaker (27.3%).

An interaction effect for speaker and noise was found, as shown by figure 5, when listening to the male speaker in the -3 dB noise condition. The observed difference between listening to the male speaker in the -3 dB SNR and listening to the male speaker in clean was significant (table 7), whereas this effect has not been found to be significant for the female speaker. This means that the effect of noise only occurs when identifying emotions in utterances by the male speaker.

Table 7: results for joy

Emotion	Variable	Estimate (β)	Std. Error	P-value
Joy	(Intercept)	.7	.71	.33
Joy	Male Speaker	-.35	.4	.38
Joy	SNR 5 dB	-.17	.4	.67
Joy	SNR -3 dB	-.66	.4	.1
Joy	Male Speaker*SNR 5 dB	-.27	.56	.62
Joy	Male Speaker*SNR -3 dB	-1.21	.59	.04

Figure 5: interaction effect of speaker and noise for joy



3.5 Neutral

Figure 1a shows that neutral was one of the best recognised emotions overall, as well as in noise. Only a small decrease is shown in the recognition rate in the -3dB listening condition compared to the clean condition. However, figure 1b shows that the false positive rate of neutral was very high, with an increase as the amount of noise increased, which means that it was very often incorrectly chosen. Figure 2 shows that neutral was recognised exactly as well in normal utterances as in nonsense utterances (all 44,8%), and figure 3 shows that neutral was recognised exactly as well in utterances by female speaker as in utterances by the male speaker (all 44,8%).

For neutral, I analysed the results the same way as the other emotions (with a generalized linear mixed effect model). However, when analysing the results, the models could not converge. Therefore, the best fitting model could not be found and the data could not be analysed, so only descriptive statistics were given.

4. Discussion

In my BA thesis I aimed to answer the following questions: in how far are Dutch people able to identify emotions in Italian; and what is the effect of noise on non-native emotion perception? Cross-linguistic emotion perception in general was investigated by studying verbal emotion perception by Dutch people in Italian. Furthermore, I investigated the hypothesis that the perception of emotions would become more difficult as background noise became more prominent and that noise would influence the emotions differentially.

For each emotion, the recognition rate was calculated by calculating how often an emotion was correctly judged as being that emotion. Moreover, noise was added as a fixed factor in the analysis to analyse the effect of noise on the recognition of the different emotions Furthermore, sentence type, speaker and trial were added as fixed factors in the analysis, to be able to account for possible influences of these factors. False positives were calculated as well, however, this was only done to be able to get a better image of what the recognition rates mean and thus were not taken into account in the analysis.

The results showed that Dutch people are able to recognise emotions in Italian speech when they do not speak Italian. Furthermore, the results showed a negative effect of noise on the perception of emotions, but only for anger, fear and joy. This means that noise does affect the perception of emotions and that this effect differs per emotion, which is according to the hypothesis. For the remaining factors, an effect of sentence type was found for anger. Also, an effect of speaker was found for anger, fear and joy. Last, an effect of trial was found for sadness.

4.1 Verbal emotion perception in clean

When listening in clean, all emotions were recognised correctly over half of the time. Anger was the best recognised emotion in clean, compared to the other emotions. The proportion of false positives for anger was below 10 percent. This shows that anger was not only recognised very well, but also that other emotions were confused with anger very few times. This indicates that anger is easily recognised, probably because it has different prosody from the other emotions and therefore is not very often confused with the other emotions. These findings are in line with the studies in which the EMOVO Corpus was validated, which showed that anger was the best recognised emotion out of the emotions in the EMOVO Corpus (Constantini et al., 2014; Giovannella et al. 2009; 2012).

Other studies, which focused on different languages, showed that anger is an easily recognised emotion in environments without background noise (e.g. Scherer et al., 2001; Thompson & Balkwill, 2006; Pell et al., 2009). The results of the current study, in addition to previous findings, give evidence that anger is one of most easily recognised emotions in general. One explanation might be the fact that anger manifests itself in the speech signal through i.e. increased loudness (e.g., Sobin & Alpert, 1999). Speaking with a louder voice, or even screaming, is one important characteristic of anger in acoustic signals. Furthermore, it is distinctive and therefore an aspect of anger that probably makes it easier to recognise compared to other emotions.

Neutral was also recognised well in the clean condition. This is in line with earlier studies (e.g. Constantini et al., 2014; Giovannella et al. 2009; 2012; Scherer et al., 2001). However, the false positive rate for neutral was nearly 40 percent. This means that participants very often confused another emotion with neutral. Some participants stated that they chose neutral as an answer when they were not sure which emotion they heard. So, neutral seems to have been used as an "I don't know" option, at least by some of the listeners. It is likely that most participants judged emotions as being neutral when they did not know which emotion it was, since they were obligated to choose one of the five options. To be able to determine the true recognition rate of neutral, an "I don't know" option could be added in a follow-up study. Thus, neutral (or any other emotion) will not function as participants' choice anymore if they are unsure.

Fear, joy and sadness were all recognised a little below 60 percent of the time. However, the false positive rate for joy was 15 percent, whereas the rates for fear and sadness lay around 35 percent. This indicates that participants less often incorrectly chose joy as an answer than fear and sadness, and that joy thus may sound more unique or distinctive than fear and sadness, despite the comparable recognition rates. Previous studies focusing on cross-linguistic emotion perception reported various recognition rates for joy (ranging from 88% for Spanish listeners in English (Pell et al., 2009), to 28% for German listeners in Indonesian (Scherer et al., 2001)) (e.g., Scherer et al., 2001; Thompson & Balkwill, 2006; Pell et al., 2009). Since the recognition rates are quite variable, it seems that the ability to identify joy is dependent on the native language of the listener and the language in which they have to identify an emotion. This could mean that joy has more language- and cultural-specific characteristics which are used by listeners to decode joy, and which might make it more difficult for listeners from other cultures to perceive joy. A study focusing on the perception of joy in different languages by listeners with different mother tongues might clarify more about the possible language-specific aspects of joy.

For fear the recognition rates reported by earlier studies on cross-linguistic verbal emotion perception vary as well, (ranging from 86% for Spanish listeners in German (Pell et al., 2009) to approximately 33% for English listeners in Japanese (Thompson & Balkwill, 2006)) but overall they are a little better than those of joy, since most recognition rates reported for fear were higher than those of joy (e.g., Scherer et al., 2001; Thompson & Balkwill, 2006; Pell et al., 2009). The fact that previous reported recognition rates, in addition to the one found in the current study, vary, points towards the possibility that fear (as well as joy) contains more language-specific characteristics than the other emotions. Therefore,

listeners from different cultures will be better or worse in judging fear when portrayed in an unfamiliar language and thus the recognition rates will differ per group of listeners.

Last, earlier studies on cross-linguistic verbal emotion perception reported overall high recognition rates for sadness (around 80%, e.g., Scherer et al., 2001; Thompson & Balkwill, 2006; Pell et al., 2009), which does not support the findings in the current study. However, previous studies did not focus on Italian and in the current study an effect of trial was found, meaning that participants recognised the items portraying sadness better as they progressed in the experimental task. This could mean that participants found it difficult to recognise sadness in Italian at first, because the language was unfamiliar, but as they judged more utterances they were able to correctly identify sadness more often. The results show a slight increase in the recognition rate for sadness from the clean condition to the 5 dB SNR condition, as well as a decrease in the false positive rate from the clean condition to the 5 dB SNR condition, but a decrease in the recognition rate from the clean condition to the -3 dB SNR condition. Therefore the overall recognition rate might have been higher if more items were judged without the interference of noise, even though the effect of noise has not been found to be significant for sadness. It may be that the set of items used in the current experiment was just too small. This could be investigated in a follow-up study, by letting participants judge more utterances portraying sadness in a clean condition (and perhaps, in noise conditions as well, to compare with the current results) than were judged in the current study.

4.2 Verbal emotion perception in noise

Studies in which verbal emotion perception by computers was investigated (e.g., Schuller et al., 2006; Zhao et al, 2013) found a negative effect of noise, meaning perception got more difficult when identifying verbal emotions in noise. An effect of noise was found in the current study as well, by focusing on verbal emotion perception by human listeners. The effect was found for the perception of anger, fear and joy, which supports the hypothesis that noise would negatively affect the perception of emotions by non-natives. Furthermore, the observed effects differ per emotion; in addition to effects of noise, effects of sentence type and speaker have been found for anger; an effect of speaker has been found for fear; and an interaction effect of speaker and noise has been found for joy. Thus, the effect of noise depends on emotion (and speaker), which is in line with the hypothesis that the effects of noise would differ per emotion.

The noise conditions were purposely designed to represent different communicational environments. The 5 dB SNR condition represented an environment in which the noise is only distracting, whereas the -3 dB SNR condition was designed to really complicate the

perception process. Significant effects have only been found for the -3 dB condition, which tells us that (as expected) severe background noise makes it difficult to perceive emotions.

A negative effect of noise was found for anger (in the -3 dB condition), meaning that the perception of anger gets more difficult as background noise gets more severe, compared to listening without background noise. However, this does not mean that anger is recognised poorly when listening in a noisy environment, since anger was the best recognised emotion in every listening condition (compared to joy, fear and sadness) in the current study, despite the negative effect of noise. This finding is supported by findings of the previous studies that focused on emotion perception in noisy environments (with a computational approach) as they found that anger was the best recognised emotion when perceiving emotions in noisy environments (Schuller et al., 2006; Zhao et al., 2013). Both the findings in the current study as well as previous findings also support the idea that anger is a very distinctive and easily recognisable emotion, since it is very well recognised in different kinds of listening conditions. Perceiving anger gets more difficult in a noisy environment compared to environments with less or without background noise, but will probably still be recognised better than other emotions in noisy environments.

Listening in the -3dB noise condition negatively affected the perception of fear, meaning that the perception of fear gets more difficult when listening in a very noisy environment. Schuller and colleagues (2006) also found that perceiving fear in noise gets more difficult as the noise gets more severe. Since the perception of fear is more difficult in general, ultimately the effect of noise is greater for fear than for anger, because it leads to poorer perception than the perception of anger. However, the decrease in recognition rate for fear is comparable to the one of anger (both approximately 10%), meaning that the effect of noise on fear is not much stronger than the effect of noise on anger. Fear just seems to be an emotion more difficult to perceive, in clean as well as in noise.

Furthermore, an interaction effect of noise and speaker was found for joy, meaning here that the male speaker was recognised significantly worse compared to the female speaker, in the -3dB noise condition. Thus, participants found it more difficult to identify emotions in severe background noise, but only for the utterances by the male speaker. This indicates that verbal emotion perception is not only more difficult when listening with background noise, but that the recognition is also dependent on the person that is speaking at that moment, which will be discussed further in section 4.4.

In the study of Van Zyl and Hanekom (2011), they found that prosody is less influenced by the presence of noise than speech. This would explain why the perception of only a few emotions is influenced by background noise, and only in the most severe noise

condition. Since the prosody of emotions (Sobin & Alpert, 1999) differs from the prosody of normal speech, it might be that prosody is one of the distinctive features people use to identify emotions, which would still be useful if it indeed is not influenced (very much) by the presence of background noise. However, further research should be done to learn more about the role of prosody in (cross-linguistic) verbal emotion perception in noise, since the current study did not focus on specific characteristics of emotions.

4.3 Sentence type

The effect of sentence type was significant for anger. When judging nonsense utterances, participants gave significantly more right answers compared to their judgments of the semantically normal utterances. Apparently, the nonsense utterances were more easily recognised. Since the participants were all Dutch, it cannot have anything to do with the content of the utterances. An explanation could be that the actors portrayed the emotions more clearly, or exaggerated more when uttering the nonsense sentences. It could be easier to portray emotions in an utterance with normal content, even though the content has nothing to do with the emotion, because it is less abstract than a nonsense utterance. An actor would still be able to imagine a context in which they could express the emotion with that utterance, whereas that would be much more difficult or even impossible when uttering a nonsense sentence. This could lead to overacting and exaggerating the emotion in nonsense utterances, which could make it easier for a listener to identify the emotion. However, the effect of sentence type was only significant for anger, meaning further investigation should be done (in native emotion perception as well) to collect more evidence on the effect of sentence type on emotion perception.

4.4 Speaker

For each emotion, utterances by two different actors were chosen to represent the emotions. Naturally, since speakers differ, their ability to portray emotions could differ as well and therefore the recognition of the different emotions could be different per speaker. This was in fact the case in the current study, since the ability to portray emotions differed per actor, even though I chose the best actor for each emotion (Giovannella et al., 2009; 2012). The fact that actors were used means that the speech was acted, which may have led to the speech being overacted. Both factors could have influenced the recognition of the emotions.

For anger, fear and joy, a significant effect of speaker was found. For anger, participants more often answered correctly when listening to the male speaker. This means

that the male speaker was significantly better than the female speaker in portraying anger. These results are in line with the findings of earlier validations (Giovannella et al., 2009; 2012) where the male actor also showed a much better recognition rate than the female actor (82% and 53%, respectively). Furthermore, participants commented that one of the male speakers (which indeed was the one selected to represent anger) sometimes screamed a bit, compared to the other speakers. This of course is a characteristic of anger as an emotion (e.g., Sobin and Alpert, 1999), and therefore not a problem, but it would explain why participants recognised anger better for the male speaker.

The results for fear showed that participants more often answered correctly when listening to the utterances of the female speaker, meaning she was better in portraying fear than the male actor. Following the findings of earlier validations (Giovannella et al., 2009; 2012), the female actor showed a much better recognition rate than the male actor (100% and 65%, respectively). Due to the very high recognition rate of the female speaker, it is clear that is the reason why her utterances were recognised so well, especially compared to the utterances of the male speaker.

Last, the results for joy showed that the utterances by the male speaker are recognised less well in the -3dB noise condition than the utterances by the female speaker, which was already shortly pointed out in section 4.2. The explanation for this finding could be the same as for fear: in table 1 the recognition rates per speaker are shown, and the female speaker is recognised much better (73%) than the male speaker (23%). Since the female actor is better in portraying joy, it is very likely that the effect of noise only appeared when identifying utterances by the male speaker because of his very low overall recognition rate for joy. However, the male speaker used to portray joy in the current experiment was the one with the best recognition rate available in the corpus, which is one of the possible limitations of using existing material. In further research, a male speaker with a greater ability to portray joy should be used to be able to study the recognition of joy in noise.

Since a significant effect of speaker was found more than once, in further research it could be considered to only use one actor per emotion. If some actors are recognised so much better than others, using the less skilled actors could negatively influence the recognition of an emotion because the actor did not portray the emotion well enough. For this reason, a validation study of an actor's ability to portray emotions should always be done, otherwise it will not be clear which actor is the better one and if the actor chosen to represent an emotion is able to represent one or more emotions. If you do use different actors, it is very important that they are able to portray emotions at approximately the same level of skill. This would mean that the actor portraying anger should be as good in portraying anger as the actor

portraying fear is in portraying fear, and so on. However, one could also argue that actors who were recognised best were recognised so well because of the fact that they are actors and the speech is acted. Because acted speech could be overacted (Wilting et al., 2006), it could be that the utterances by the 'best' actors are recognised this well because of overacted speech. This hypothesis could be tested by using natural speech in future research, in which emotions are natural rather than acted. Using natural speech may also cause different effects from different speakers, because speakers differ in the way they portray emotions, but as the speech is no longer acted, any effect of overacting is excluded and effects in speaker performance are caused by natural circumstances.

5. Conclusion

Dutch listeners are able to recognise emotions in Italian, without having any knowledge of the language. This is in line with previous findings that verbal emotion perception can happen cross-culturally. Furthermore, they are able to identify emotions in noisy environments as well as in environments without background noise. However, for anger, joy and fear a negative effect of noise has been found for the condition with strong background noise, meaning that perceiving these emotions gets more difficult when listening in an environment with severe background noise. Nevertheless, the effect has only been found for three emotions, of which anger is the best recognised emotion overall (despite the effect of noise). This means that emotions are still quite distinguishable in background noise, and for this reason, in future research more difficult SNRs should be used as well, to see when the other emotions will become more difficult to perceive. Furthermore, since the prosody of emotions differs from the prosody of speech, another option to learn more about verbal emotion perception may be to study which characteristics of specific emotions (i.e., prosody) are the ones that 'survive' background noise. This will tell us more about the role of prosody in and emotion-specific characteristics used to identify verbal emotions.

For joy, the effect of noise appeared in an interaction with speaker, meaning it only appeared when participants identified emotions in utterances from the male speaker. For fear and anger, an effect of speaker was found as well. This indicates that not only the type of noise may interfere with verbal emotion perception, but that the ability to perceive emotions is also dependent on the person portraying an emotion. For the current study, the effect is probably due to the fact that actors were used to portray the utterances. Not all actors were equally skilled to perform emotional speech, which most likely led to variety in the recognition of emotions based on which speaker portrayed them. Therefore, in future research the skills of the speakers should not only be taken into account, but they should be controlled

and their ability to portray emotions should be comparable, also to avoid overacting. Note, however, that it is likely that the way speakers portray emotions in natural circumstances differs as well. Another option may be using natural speech to study verbal emotion perception (in noise, or in general). In this way, if any differences are found between speaker performances, it is certain that these effects have been caused by natural circumstances rather than the skills of an actor.

Furthermore, even though the participants did not speak Italian, an effect of sentence type has been found for anger. In the current study this is most likely caused by overacting. However, this is not further investigated. Therefore, sentence type should be taken into account when testing natives as well, to be able to study in how far nonsense or normal utterances influence the perception of emotions.

References:

Audacity Team, 2016. Retrieved from: www.audacityteam.org.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. 59, 390–412.

Broersma, M., Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication* 52, 2010, page 980-995.

Constantini, G., Iadarola, I., Paoloni, A., & Todisco, M. (2014). EMOVO Corpus: an Italian Emotional Speech Database. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, LREC, 2014.

Coumans, J., van Hout, R. & Scharenborg, O. (2014). Non-native Word Recognition in Noise: The Role of Word-initial and Word-final Information. *Interspeech* 2014, Singapore, 14-18 September.

Da Silva, W., Barbosa, P.A. & Abelin, Å. (2016). Cross-cultural and cross-linguistic perception of authentic emotions through speech: An acoustic–phonetic study with Brazilian and Swedish listeners. *Delta*, vol. 32, no. 2, Sao Paulo, May/August 2016.

Darwin, C. (1872). The expression of the emotions in man and animals. London: Murray.

Ekman, P. & Friesen, W.V. (1971). Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*, 1971, Vol. 17, No. 2, page 124-129.

Ekman, P. (1992). An argument for Basic Emotions. *Cognition and Emotion*, 1992, 6, page 169-200.

Ekman, P. (1999). Basic Emotions. *Handbook of Cognition and Emotion*, 1999, Chapter 3.

Giovannella, C., Conflitti, D., & Santoboni, R. (2009). Transmission of vocal emotion: Do we have to care about the listener? The case of the Italian speech corpus EMOVO. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, Amsterdam, 2009, pp. 1-6.

Giovannella, C., Floris, D., & Paoloni, A. (2012). An Exploration on Possible Correlations among Perception and Physical Characteristics of EMOVO Emotional Portrayals. *Interaction Design and Architecture(s) Journal*, 10, pp. 102-111.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, page 288-299.

Kerkhoff, J. (2015). Audio Manipulation. *Praat Script*, University of Nijmegen, Department of Language and Speech.

Pell, M.D., Monetta, L. & Paulmann, S. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behaviour,* 2009, Vol. 33, page 107-120.

R development core team (2011). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna.

Scharenborg, O., Sanders, E., Cranen, B. (2014). Collecting a corpus of Dutch noise-induced 'slips of the ear'. *Proceedings of Interspeech* 2014, Singapore, 2600-2604.

Scherer, K.R., Banse, R. & Wallbott, H.G., (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-cultural Psychology,* Vol. 32 (1), page 76-92, January 2001.

Scherer, K.R., Clark-Polner, E. & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology,* 2011, 46 (6), page 401-435.

Schuller, B., Arsić, D., Wallhoff, F. & Rigoll, G. (2006). Emotion Recognition in the Noise Applying Large Acoustic Feature Sets. *Speech Prosody,* ISCA.

Sobin, C. & Alpert, M. (1999). Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sadness and Joy. *Journal of Psycholinguistic Research,* Vol. 28 (4), page 347-365.

Thompson, W.F. & Balkwill, L-L. (2006). Decoding speech prosody in five different languages. *Semiotica,* 158-1/4 (2006), page 407-424.

Van Zyl, M. & Hanekom, J.J. (2011). Speech Perception in Noise: A Comparison Between Sentence and Prosody Recognition. *Journal of Hearing Science,* 2011, Vol. 1 (2).

Wilting, J., Krahmer, E., and Swerts, M. (2006). Real vs. acted emotional speech. In *Proceedings of the 9th International Conference on Spoken Language Processing*, Pittsburgh, page 805–808.

Zhao, X., Zhang, S. & Lei, B. (2013). Robust emotion recognition in noisy speech via sparse representation. *Neural Computing and Applications,* Vol. 24 (7), page 1539-1553.