Age, Hearing Loss and the Perception of Affective Utterances in Conversational Speech

Juliane Schmidt[1,2], *Esther Janse*[1,3,4], *and Odette Scharenborg*[1,3]

[1] Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[2] IMPRS for Language Sciences, Nijmegen, The Netherlands
[3] Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, The Netherlands
[4] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

{J.Schmidt, E.Janse, O.Scharenborg}@let.ru.nl

Abstract

This study investigates whether age and/or hearing loss influence the perception of the emotion dimensions arousal (calm vs. aroused) and valence (positive vs. negative attitude) in conversational speech fragments. Specifically, this study focuses on the relationship between participants' ratings of affective speech and acoustic parameters known to be associated with arousal and valence (mean F0, intensity, and articulation rate). Ten normal-hearing younger and ten older adults with varying hearing loss were tested on two rating tasks. Stimuli consisted of short sentences taken from a corpus of conversational affective speech. In both rating tasks, participants estimated the value of the emotion dimension at hand using a 5-point scale. For arousal, higher intensity was generally associated with higher arousal in both age groups. Compared to younger participants, older participants rated the utterances as less aroused, and showed a smaller effect of intensity on their arousal ratings. For valence, higher mean F0 was associated with more negative ratings in both age groups. Generally, age group differences in rating affective utterances may not relate to age group differences in hearing loss, but rather to other differences between the age groups, as older participants' rating patterns were not associated with their individual hearing loss.

Index Terms: affective speech, age, hearing, natural speech.

1. Introduction

The speech signal does not only contain information on *what* has been said, but also contains information on *how* the speaker feels about the content. Affective prosody enhances emotional information processing [1] and prosodic acoustic cues are crucial for the correct interpretation of certain affective expressions, e.g. *irony* [2]. Affective information can be described in several ways. The categorical approach, meaning that concrete terms such as *happy*, *sad*, *neutral*, *bored*, or *angry* are used to describe different affective expressions, seems to be predominant in emotion perception research on older populations [3–5]. However, this approach has certain drawbacks as these discrete and rather static concepts do not capture emotion blends. Further, they may bias the responses of participants, or may not be consistent with the participant's own interpretation of a particular affective state [6,7]. A more detailed description of affect can be achieved by a dimensional and more continuous approach. Here, emotions are plotted into a two or three dimensional space, where the most frequently used axes are arousal (calm-aroused) and valence (negative-positive) [7]. Several concepts have been proposed as a third dimension. Examples are tension, control, potency [7], or dominance [e.g., 8]. Compared to the categorical approach, the dimensional approach disentangles the relative contribution of each dimension to emotion categories.

Affect in speech is expressed by physiological changes in articulation, resulting in a unique acoustic pattern for every type of affect expressed in speech [6,9,10]. Certain acoustic parameters have been shown to correlate with specific emotion dimensions. Mean F0, mean intensity, and speech rate have been investigated most often. For arousal, Pereira [11] reported higher mean F0 and increasing mean intensity with higher degrees of arousal for both male and female voices. This is supported by Schröder et al. [12], who further report longer phrases and shorter pauses for aroused utterances. For valence, weaker correlations with acoustic measures were reported. Moreover higher F0 for more positive utterances (for male speakers) [11], increasing intensity and longer pauses for more negative utterances [12] were reported.

Aging has been shown to result in changes in the perception of prosodic features (such as intensity, speech rate and F0), particularly of F0 [13]. Age has indeed been found to influence perception of affective information in speech [3,5,14]. For instance, Paulmann et al. [14] showed that young participants were significantly better at recognizing *anger*, *disgust*, *fear*, *happiness*, and *sadness* from prosodic information than older adults.

Moreover, hearing loss aggravates auditory processing difficulties for some of the acoustic parameters mentioned, e.g., intensity and speech rate [15]. However, as older participants in [5,13] had near-normal hearing, the role of differences in individual hearing loss in this age effect on perception of affective information remains unclear. So far, these age effects have not been attributed to hearing loss, but this might be due to several methodological reasons. First, hearing loss was either not assessed properly in those studies [5,14], or it was not related to the acoustic parameters of the stimuli [3]. Second, all three studies [3,5,14] used stimuli that were acted or artificial. Acted stimuli, however, might not reflect the acoustic details needed for a correct perception and classification of affect in everyday speech [9]. Though there are also disadvantages to natural speech material in emotion research (i.e., small number of speakers, short utterances, and poor recording quality), ecological validity is highest in natural speech stimuli [6]. Acted speech is likely to be overacted, particularly the acoustic cues for arousal [9], resulting in the use of a more prototypical acoustic expression. This may lead to a more extreme realization of affective prosody in acted speech [16]. For instance, acted *anger*, for which mean intensity would be a prominent feature [9,17], may be relatively easy to perceive if overdone, even for people

with hearing loss. In natural speech, however, affective information is cued much more subtly [9]. Taken together, these findings are not conclusive as to whether hearing loss may influence the perception of affect in natural speech.

In this study, the role of age and hearing loss in the perception of affective utterances is investigated. We combine three aspects of affective speech perception that have not been combined in the studies described above: the perception of affective utterances is investigated by using a dimensional approach, by using natural (i.e., non-acted) speech stimuli, and by linking acoustic parameters directly to an individual's hearing loss. The focus is on arousal and valence because these emotion dimensions are used most consistently across studies. The first research question that we address is whether younger and older listeners differ in the perception of affect and in the way they make use of the corresponding acoustic parameters. This is investigated by comparing the associations between acoustic parameters and the affective ratings of the two age groups. The two age groups are compared using two separate one-dimensional affective rating tasks, one for arousal and one for valence. The second research question asks whether and how differences in hearing loss among the older adults impact their affective ratings.

2. Experimental Set-up

2.1. Participants

Two groups of 10 native German participants were recruited (50% male participants in each group). The younger group consisted of students from Saarland University in Saarbrücken (age: M = 25.0; SD = 2.0; range: 22 – 28 years) and the older group was recruited from the greater area of Saarbrücken (age: M = 59.7; SD = 5.6; range 53 – 68 years). None of the participants used a hearing aid in daily life. Participants' hearing was assessed prior to testing by a professional hearing aid audiologist. Pure tone thresholds were retrieved for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz for both ears. The average of both ears was used for the analysis, as the degree of hearing loss did not differ significantly between the left and the right ear. Mann-Whitney U Test for independent samples showed that the younger group had significantly better hearing in the higher frequencies, i.e., from 1 kHz up to 8 kHz ($ps < .05$) than the older group. Individual pure-tone average (PTA) over participants' thresholds at 1, 2, and 4 kHz over both ears was entered as an index of hearing loss in the analyses below. Mean PTA for the younger listener group was 3.61 dB HL (SD = 3.16, range: 1.67 – 13.33 dB HL) and for the older listener group was 16.33 dB HL (SD = 8.98, range: 6.67 – 31.67 dB HL).

2.2. Speech material

The stimuli were taken from the audio-only section of the audio-visual "Vera am Mittag" corpus (henceforth: VAM corpus) for authentic and affectively colored conversational speech [8]. The VAM audio corpus consists of 1018 affective utterances divided into two subsets. Utterances were taken from a German TV talk show. The first subset (VAM-I) consists of 499 utterances produced by 19 different speakers (4 m/15 f). The second subset (VAM-II) consists of 519 utterances by 28 speakers (7 m/21 f). Moreover the corpus provided mean affective ratings for the degree of arousal and valence for each utterance. These affective ratings were collected by means of a pictorial 5-point-scale [18], consisting of five line drawings of a human figure. Each figure expresses a different degree of arousal or valence through changes in attributes, i.e., indication of tremble or change of facial expression. A numeric value was attached to each point, ranging in 0.5 steps from -1 (very calm/very negative) to 1 (very aroused/very positive). Each utterance was evaluated by a group of presumably younger adults ($N = 17$ for VAM-I, $N = 6$ for VAM-II, their age is not documented). Their mean ratings per stimulus are treated as ground truth in our analysis.

According to the ground truth, the VAM corpus provides a good coverage of the emotional space (arousal range: -0.83 – 1.00; valence range: -0.80 – 0.77). However, due to the discussion topics within this TV format (relationship crises, jealousy, fatherhood questions, etc.), the emphasis within the corpus was found to be on neutral to more negative emotions [8].

2.2.1. Subsets for the arousal and valence rating tasks

Stimuli were selected from both VAM-I and VAM-II. In order to not confuse or bias participants, only one emotion dimension per task was rated. Hence, we created two separate stimulus sets: one for arousal and one for valence that complied with the following three criteria. First, as the temporal window for information integration is limited [19], utterance duration had to be shorter than three seconds. Second, when interpreting an utterance, both verbal (*what* is said) and non-verbal (*how* something is said) information is used [1]. In fact, the semantic meaning may change the emotional content of an utterance, e.g., when non-verbal information is negative while the verbal information is positive, as in sarcasm [2]. As we were exclusively interested in the non-verbal information, utterances had to be as semantically neutral as possible to minimize semantic interference (e.g., *Du bist der Vater,* 'You are the father'). Utterances were therefore presented in written form to three independent evaluators who rated whether these sentences were semantically neutral. Only the utterances labeled as neutral by at least two of the three raters were included in the final stimulus sets. Third, utterances had to have a ground truth value for either arousal or valence that was close to the value of the five points on the scale (1, 0.5, 0, -0.5, -1). In order to familiarize participants with the task, four additional utterances per rating task served as practice trials.

The final item set for the arousal rating task consisted of 9 utterances from VAM-I and 15 utterances from VAM-II (17 different speakers in total; 3 m/14 f; rating range: -0.66 – 0.94). The item set for the valence rating task included 7 utterances from VAM-I and 11 utterances from VAM-II (15 different speakers in total; 4 m/11 f; rating range: -0.80 – 0.77). There was an overlap of two utterances between the item sets; hence, two stimuli were rated for both dimensions.

2.3. Procedure

Ratings of degree of arousal and valence by a group of younger and a group older adults were collected with a simple pen-and-paper version of the pictorial rating tool that was used in the VAM corpus [18]: five line drawings of a figure depicted five states along the dimension of either arousal (calm – expressive) or valence (frowning – smiling).

Prior to each rating task, the emotion dimension at hand was explained to the participant. Further, the pictorial rating tool was introduced by describing the meaning of each point on the scale. Participants' attention was particularly drawn to

the changing attributes of the figure. Finally, it was emphasized that listeners could give only one rating per trial and that they should indicate their choice by marking the figure on the scale. Additionally, written instructions were provided and there was the possibility to ask questions.

Both age groups completed the arousal rating task first, followed by the valence rating task, with a short break in between the two tasks. Participants were seated in a sound-attenuated booth and heard the utterances via closed headphones connected to a laptop. Each utterance was presented twice, i.e. as two separate trials. The order in which stimuli were presented was randomized for each participant, using the experimental program SCAPE [20]. Participants would always start with the practice trials. The utterances were played at the same fixed volume to both age groups. Listeners were allowed to listen to each trial several times before making a decision. For the arousal rating task, 29% of the utterances were listened to more than once, and for the valence ratings task, 20% of the utterances were repeated. Tasks were completed in the participant's own pace but completing both tasks did not take more than 30 minutes.

3. Results

First, we will report the acoustic measurements of the stimuli for both tasks and their relation to the ground truth to demonstrate that the acoustics were related to the two emotion dimensions. Second, to compare the age groups' ratings, we used linear mixed effects regression analyses. As each utterance was rated twice by the participants, we calculated the average rating per stimulus. The initial model allowed for interactions between age group and all acoustic parameters (with stimulus and participant as random effects). We arrived at the best fitting model by a stepwise exclusion of interactions and predictors with the highest non-significant *p*-values. Next, we investigated whether individuals' hearing loss was associated with their rating of affective information. Therefore, we checked for interactions between hearing loss and the acoustic parameters in the older adults' data. The model fitting procedure was identical to the one used in the group comparison.

Table 1. *Pearson correlation coefficients between acoustic predictors and reference ratings for arousal and valence.*

		Mean Intensity	Articulation Rate	VAM Value
Arousal	Mean F0	.80 ***	-.38	.71 ***
	Mean Intensity	–	-.42*	.91 ***
	Articulation Rate	–	–	-.38
Valence	Mean F0	.67***	-.16	-.35
	Mean Intensity	–	-.21	.06
	Articulation Rate	–	–	.20

* $p < .05$, ***$p < .001$

3.1. Acoustic measurements section

Mean F0 and mean intensity for each utterance were measured using Praat [21]. Articulation rate was calculated by dividing the number of syllables by file length minus pauses (i.e., pauses longer than 100 ms).

For arousal, we found strong positive correlations for mean F0 and mean intensity with the ground truth (VAM Value). Hence, higher mean intensity and higher mean F0 is associated with higher levels of arousal. The correlation between articulation rate and the ground truth was not significant. Moreover, there were no significant correlations between the ground truth for valence (VAM Value) and any of the acoustic parameters (see Table 1).

3.2. Rating analysis

Figure 1 shows participants' ratings (y-axis) compared to the ground truth (x-axis). Star symbols show the mean of the younger participants' ratings for each individual utterance, and triangles show the mean of the older participants' ratings for each individual utterance. Fit lines have been added, where the dashed line depicts the fit line through the younger participants' ratings, and the solid line depicts the fit line through the older participants' ratings. Figure 1a shows the results for arousal, and Figure 1b for valence.

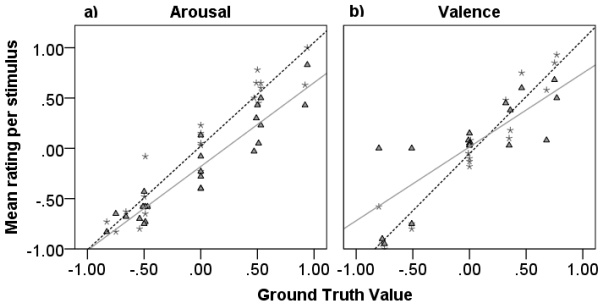

Figure 1: *Mean ratings per utterance of the younger (star symbols, dashed lines) and older participants (triangle symbols, solid lines) as a function of ground truth values.*

Table 2. *Final models of the linear mixed-effects regression rating analyses.*

		Age Group Comparison		Older Adults	
		B	*t*	*β*	*t*
Arousal	Group	-.20 **	-3.27	–	–
	Mean intensity	.11 ***	10.49	.09 ***	7.90
	Group*Mean Intensity	-.03 ***	-3.95	–	–
Valence	Group	.06	1.03	–	–
	Mean intensity	.08	1.70	–	–
	Mean F0	-.01*	-2.41	-.00 **	-3.42
	Group*Mean Intensity	-.06 ***	-4.87	–	–

* $p < .05$, ** $p < .01$, ***$p < .001$

3.2.1. Analysis of arousal

Figure 1a shows that younger participants are more in agreement with the ground truth than the older participants, as the latter group diverges more from the diagonal. We carried out a statistical analysis to investigate whether this age group difference is significant, and to investigate the influence of the acoustic parameters on the ratings. The Arousal panel in Table 2 displays the best fitting models for both the age group comparison and the separate analysis of the older adults group.

The statistical analysis revealed significant effects of mean intensity and age group on perceived arousal and an age group by mean intensity interaction. This implies that the higher the stimulus' mean intensity was, the more aroused younger participants (mapped on the intercept) rated the utterance. Older adults generally rated the utterances as less aroused, and showed a smaller effect of intensity on their ratings.

As in the age group comparison, the analysis of the older adults' data only showed an effect of intensity on perceived arousal. Thus, higher intensity is perceived as more aroused among the older group. There was no effect of hearing loss.

3.2.2. Analysis of valence

Figure 1b shows that stimuli on both ends of the ground truth scale are rated as less extreme by older participants compared to younger participants. To investigate this, we carried out a similar statistical analysis as for arousal. The best fitting models are displayed in the Valence panel in Table 2.

The age group comparison for valence showed a significant effect of mean F0 and a significant interaction between age group and mean intensity. Hence, higher mean F0 led to more negative ratings in both age groups. While younger adults (mapped on the intercept) do not use mean intensity to rate valence, older adults seem to interpret increasing intensity as more negative.

Analogous to the analysis of the arousal task data, the data of the older adults were analyzed separately to investigate the effect of hearing loss. To that end, hearing loss was allowed to interact with the acoustic predictors. Results showed a significant effect of mean F0: higher mean F0 yields more negative ratings of valence. Though older adults seem to use intensity differently than younger adults in the age group analysis above, the effect of mean intensity is not strong enough to surface in the subset analysis here. Moreover, hearing loss did not affect valence ratings, nor did it interact with the acoustic parameters.

4. General Discussion

This study was set up to investigate possible age effects in the perception of affect in speech in relation to several acoustic parameters of the stimulus. In contrast to earlier studies on age differences in affect perception, we combined three methodological aspects to investigate the perception of emotional content, i.e., linking the perception of acoustic parameters with individual hearing loss, conversational (rather than acted) speech, and a rating of emotional dimensions, rather than categorical classification.

The expression of affect has been related to acoustic parameters, which listeners use to interpret affect [10]. In our study, we investigated the relation between acoustic parameters and affective ratings in two ways. First, we investigated how the acoustic parameters in our stimuli correlated with the ground truth affective ratings that came with the corpus materials. Second, these acoustic parameters were entered as predictors of participants' ratings of the conversational stimuli.

Arousal stimuli showed positive correlations between the ground truth arousal ratings from the corpus and both mean F0 and mean intensity. These correlations are comparable to correlations reported in the literature [11,12]. The association between intensity and arousal is further supported by participants' ratings. Our results showed that older adults perceived aroused utterances as less aroused than younger adults, and the effect of intensity on arousal ratings was smaller compared to its effect on younger adults' ratings. This finding agrees with the results of previous studies. Paulmann et al. [14], for instance, found that older adults classified a stimulus more often as *sad* when *fear* was the intended emotion and more often as *happy* when *pleasant surprise* was the intended emotion. If the positions of these emotions on the arousal axis are considered (e.g., [7]), older adults more often choose the term that is linked to the less aroused emotion (*sad*, *happy*) while younger adults prefer the term that is related to the more aroused term (*fear*, *pleasant surprise*).

For valence, there were no significant correlations between the ground truth and the acoustic measures investigated in this study. This is not surprising, given that correlations between valence and acoustic parameters found in larger item sets were also less strong as compared to arousal [11,12]. In line with the ground truth evaluator panel (see correlations with VAM ratings in Table 1), we did not find any evidence that the younger adults in our study based their valence ratings on mean intensity, nor on articulation rate. Older adults, however, seem to make more use of mean intensity when rating valence than younger adults though the intensity effect does not reach significance in the subset analysis of the data of the older adults. In addition, our data suggests that mean F0 plays a crucial role for both age groups when rating valence. In other words, independent of age, higher mean F0 is associated with more negative utterances. This result is in opposition to the finding of Pereira [11], but note that her result held for male talkers only (the majority of our talkers being female).

Our second question concerned the impact of hearing differences among the older adults on their ratings of affect. All in all, age effects on affect perception seemed to outweigh hearing effects. Older age led to smaller-sized intensity effects in the affect dimension. However, the older adults' data do not provide any indication that hearing loss may account for this age effect. This is in line with the findings of Orbelo et al. [3] who also did not find that hearing measures predicted affective ratings. Note, that our sample size was small (10 participants per age group) and participants' hearing was still rather good. In order to better account for possible effects of hearing loss, future work should include more participants representing a broader range in hearing loss.

5. Conclusion

Taking a dimensional approach, we tried to link acoustic variation as found in natural conversational speech stimuli to ratings of affect in younger and older adults. Our results show age effects on affect perception, and show that older age led to smaller-sized effects of acoustic differences on affective ratings. No effect of hearing loss on affective rating was observed in this older adult sample. Future research should aim at including more participants covering a broader range of hearing loss to obtain a better picture of how hearing loss may influence perception of affect in conversational speech.

6. Acknowledgments

The data for this study was collected for the MA thesis of J.S. in 2011 at Saarland University Saarbrücken, Germany, and re-analyzed in 2014. J.S.'s research was supported by the European Commission (FP7-PEOPLE-2011-290000). The research by E.J. and O.S. was supported by two individual Vidi-grants from NWO.

7. References

[1] M. D. Pell, A. Jaywant, L. Monetta, and S. a Kotz, "Emotional speech processing: disentangling the effects of prosody and semantic cues," *Cogn. Emot.*, vol. 25, no. 5, pp. 834–853, Aug. 2011.

[2] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Commun.*, vol. 50, no. 5, pp. 366–381, May 2008.

[3] D. M. Orbelo, M. A. Grim, R. E. Talbott, and E. D. Ross, "Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline," *J. Geriatr. Psychiatry Neurol.*, vol. 18, no. 1, pp. 25–32, Mar. 2005.

[4] D. M. Isaacowitz, C. E. Löckenhoff, R. D. Lane, R. Wright, L. Sechrest, R. Riedel, and P. T. Costa, "Age differences in recognition of emotion in lexical stimuli and facial expressions.," *Psychol. Aging*, vol. 22, no. 1, pp. 147–59, Mar. 2007.

[5] I. Kiss and T. Ennis, "Age-related decline in perception of prosodic affect," *Appl. Neuropsychol.*, vol. 8, no. 4, pp. 251–254, Jan. 2001.

[6] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, no. 1–2, pp. 227–256, Apr. 2003.

[7] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, Dec. 2005.

[8] M. Grimm, K. Kroschel, and S. Narayanan, "The Vera am Mittag German audio-visual emotional speech database," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2008, pp. 865–868.

[9] K. R. Scherer, "Vocal affect expression: a review and a model for future research.," *Psychol. Bull.*, vol. 99, no. 2, pp. 143–165, Mar. 1986.

[10] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *J. Pers. Soc. Psychol.*, vol. 70, no. 3, pp. 614–636, Mar. 1996.

[11] C. Pereira, "Dimensions of emotional meaning in speech," in *Proceedings of the ISCA Workshop on Speech and Emotion*, 2000, pp. 25–28.

[12] M. Schröder, R. Cowie, E. Douglas-Cowie, M. Westerdijk, and S. Gielen, "Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis," in *Proceedings of the 7th European Conference on Speech Communication and Technology (EUROSPEECH '01)*, 2001, pp. 87–90.

[13] P. Souza, K. Arehart, C. W. Miller, and R. K. Muralimanohar, "Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing," *Ear Hear.*, vol. 32, no. 1, pp. 75–83, Feb. 2011.

[14] S. Paulmann, M. D. Pell, and S. A. Kotz, "How aging affects the recognition of emotional speech.," *Brain Lang.*, vol. 104, no. 3, pp. 262–269, Mar. 2008.

[15] L. C. Cox, S. L. McCoy, P. A. Tun, and A. Wingfield, "Monotic auditory processing disorder tests in the older adult population," *J. Am. Acad. Audiol.*, vol. 19, no. 4, pp. 293–308, Apr. 2008.

[16] J. Wilting, E. Krahmer, and M. Swerts, "Real vs. acted emotional speech," in *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech)*, 2006, pp. 805–808.

[17] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: different channels, same code?," *Psychol. Bull.*, vol. 129, no. 5, pp. 770–814, Sep. 2003.

[18] P. J. Lang, "Behavioral treatment and bio-behavioral assessment: computer applications," in *Technology in mental health care delivery systems*, J. B. Sidowski, J. H. Johnson, and T. A. Williams, Eds. Norwood, NJ, 1980, pp. 119–137.

[19] E. Pöppel, "Lost in time: a historical frame, elementary processing units and the 3-second window," *Acta Neurobiol. Exp. (Wars).*, vol. 64, no. 3, pp. 295–301, Jan. 2004.

[20] R. Grabowski and D. Bauer, "System for Computer-Aided Perception Experiments (SCAPE)." 2004.

[21] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer." 2011.