



Clear Speech - Mere Speech? How segmental and prosodic speech reduction shape the impression that speakers create on listeners

Oliver Niebuhr

Mads Clausen Institute, University of Southern Denmark, Sønderborg, Denmark

oniebuhr@mci.sdu.dk

Abstract

Research on speech reduction is primarily concerned with analyzing, modeling, explaining, and, ultimately, predicting phonetic variation. That is, the focus is on the speech signal itself. The present paper adds a little side note to this fundamental line of research by addressing the question whether variation in the degree of reduction also has a systematic effect on the attributes we ascribe to the speaker who produces the speech signal. A perception experiment was carried out for German in which 46 listeners judged whether or not speakers showing 3 different combinations of segmental and prosodic reduction levels (unreduced, moderately reduced, strongly reduced) are appropriately described by 13 physical, social, and cognitive attributes. The experiment shows that clear speech is not mere speech, and less clear speech is not just reduced either. Rather, results revealed a complex interplay of reduction levels and perceived speaker attributes in which moderate reduction can make a better impression on listeners than no reduction. In addition to its relevance in reduction models and theories, this interplay is instructive for various fields of speech application from social robotics to charisma coaching.

Index Terms: speech reduction, personality traits, perception.

1. Introduction

Speech reduction in the sense of shorter, fewer, and smaller movements in speech production has been a dynamic research topic ever since cross-linguistic acoustic and articulatory evidence blurred the dividing line that had been drawn by theoretical linguists between phonetic coarticulation on the one hand and phonological processes like assimilation, lenition, and deletion on the other [1]. According to this dividing line, coarticulation was supposed to be gradual and, due to the biomechanical sources of the variation, also language-universal. In contrast, phonological processes were assumed to be pre-planned, categorically executed, and language-specific.

However, the production and perception experiments of [2,3,5] showed that both deletions and assimilations of place of articulation are gradual and in fact virtually never complete. They leave phonetic residuals in the surrounding segments. These residuals can be so small that they are hardly measurable in the acoustic signal. Sometimes, they take the shape of articulatory prosodies that are spread out over a larger stretch of speech [6,7]. Yet, listeners can pick them up and use them for lexical identification and, in the case of regressive assimilation, also for identifying the phonetic features of the upcoming sound segment that caused the assimilation [5]. Studies like those of [8] and [9] added to the picture that coarticulation patterns differ across languages and can be pre-planned. Thus, they cannot simply result from biomechanical interactions between the active and passive articulators of the

speech production apparatus. As a consequence of these discoveries, researchers started using terms like assimilation and coarticulation synonymously [10] and concluded that the postulated disparity between linguistic and biomechanical phenomena "has not been objectively justified" [11:35].

The possibilities of analyzing huge speech corpora in the new world of "digital phonetics" have further enriched our understanding of reduction by four major insights: Firstly, reduction is a much more pervasive phenomenon than has often been assumed. Spontaneous-speech data revealed for quite a few words and even entire word classes in different languages that unreduced (i.e. full) citation forms do not represent "canonical forms" insofar as they occur less often than their reduced variants [12,13]. For [13], the discrepancy between the canonical target state of words and their actual realization in spontaneous speech is so large, that they "appear to be two different levels that are only loosely connected".

Secondly, reduction has a considerable impact on canonical word forms. While reduction in read speech can typically still be described as a modification or deletion of individual sound segments, reduction in spontaneous speech often affects entire syllables or words can make them "disappear" - at least in terms of their linear segmental manifestation [6,7,12,14].

Thirdly, the degree of reduction is to some extent predictable and varies with basic linguistic and cognitive parameters. These include speaking rate, speaking style (read vs. spontaneous speech), position in the prosodic phrase, frequency of the corresponding or a neighboring (lexical) item, semantic predictability (including multi-word expressions), lexical stress, information value (given vs. new information), and co-occurrence with pitch accents [12,13,15,16].

Fourthly, qualifying the third insight, reduction is "not a fully automatic process" either [12]. Or, in the words of [13]: "the portion of unpredicted cases is still significant". For example, [12] found many instances of Dutch "eigenlijk" that were pitch-accented and still realized highly reduced.

The partial contradiction between insights 3 and 4 made [17,18,20] look for additional factors that can explain the variation in speech communication. They concluded on the basis of own and external experimental evidence that the degree of reduction also varies as a function of conveying communicative meanings. For example, sincere utterances differ from their sarcastic counterparts in that the latter show a higher degree of reduction (higher frequencies of assimilation, lenition, and deletion). Utterances conveying disagreement can be less reduced than utterances signaling agreement [21].

Furthermore, variation in the degree of reduction also tells the listener something about the speaker him/herself. Female speakers seem to reduce more than male speakers, and young speakers more than old speakers [22,23]. The production experiment of [24] showed in addition that reduction varies with

the speaker's emotional state. Compared to emotionally neutral utterances, fear and sadness caused higher and anger lower degrees of reduction.

Continuing this line of research on the links between speech reduction levels and personal characteristics, the present paper is concerned with speaker attributes. The aim is to find out whether clear speech is mere speech, or, in other words, whether general changes in reduction level are, under otherwise constant communicative conditions, systematically associated with changes in speaker attributes. In an initial attempt to pursue this aim, a perception experiment was carried out here, taking into account three different types of speaker attributes. They cover key aspects of the Big Five personality traits summarized in [4,30]: physical attributes like 'athletic' and 'clumsy', social attributes like 'sociable' and 'vain', and cognitive attributes like 'sincere' and 'scatty'.

Determining if and which speaker attributes change due to differences in reduction level, which parameters of the speech signal are involved in inducing these changes, and how strong the reduction has to be for each of these parameters are important questions to be addressed. Answers to these questions can, firstly, inform future speech-corpus analyses, in this way potentially helping us explain more (primarily cross-speaker) variation in speech reduction. Secondly, answering the above questions will be useful in compiling speaker samples for production or perception experiments; and they will contribute to refining or revising the simple tug-of-war mechanism in Lindblom's Hypo-Hyper (H&H) theory [25,20,15]. Thirdly, gaining a better understanding about how speaker attributes emerge in listeners is also of growing interest in the more applied and rapidly growing areas of persuasive technology [26,27], social robotics [28,29], and leadership or charisma training [31,32,33]. For example, knowing more about how speech production and speaker attributes are linked would allow engineers to adjust speech-synthesis algorithms such that their output supports the physical, social, and cognitive attributes of the robots in which they are integrated. Communication and leadership trainers, i.e. so-called "speech coaches", would be able to better instruct their clients what to do or avoid in speech production in order to make the desired impression on their audience.

Based on the definition of speech reduction as shorter, fewer, and smaller movements in speech production, the conducted perception experiment went beyond reduction in the segmental domain and include - for the first time as far as we are aware - reduction in the *prosodic* domain. Three levels of segmental and prosodic reduction were created and cross-combined in the stimuli of the perception experiment. The language of the experiment is German. All further details of the experimental method are explained in the following.

2. Method

2.1. Participants

Forty-six listeners took part in the experiment, 28 females and 18 males. They were between 21-46 years old (average age 27.3 years). None of them was a professional speaker or a student of linguistics or phonetics. In fact, most participants were already pursuing a professional career outside academia when they were recruited for the experiment. They received a small reward (a little treat) for their participation. All participants reported normal hearing abilities and had German as their native language.

2.2. Stimuli

The stimuli were based on the on the largely sonorant and semantically/pragmatically neutral statement "Maren fährt mit dem Wagen von Emden nach Bremen und zurück" (Maren drives with the car from Emden to Bremen and back).

The statement was realized naturally by 9 advanced female Master students of phonetics with 3 levels of reduction: (1) *unreduced*, (2) *moderately reduced*, and (3) *strongly reduced*. Unreduced means that the statement was realized fluently as a single prosodic phrase but with each word pronounced according to its officially correct (full) form specified in the DUDEN, i.e. the major pronunciation dictionary for German, published in close collaboration with the national Institute for the German Language [34]. The moderate reduction level implemented the two main types of reduction in German, i.e. schwa elision and function-word reduction, the latter to the extent that is typical of spontaneous speech [35]. The strong reduction level additionally included considerable reductions of all content words in the statement, obeying the regularities and limits of reduction that have been empirically defined by [35]. Note that the unreduced and the strongly reduced renderings of the target sentence can both occur in German and do not sound odd in the ears of German listeners. From the point of view of everyday (spontaneous) conversation, the unreduced rendering represents careful hyper-articulated speech, whereas the strongly reduced rendering would probably be considered sloppy, hypo-articulated speech.

Table 1: 3 degrees of segmental reduction in the stimuli: unreduced (U), moderately reduced (M), strongly reduced (S).

	"Maren fährt mit dem Wagen von Emden nach Bremen und zurück"
Unreduced	ma:ʁən feətʰ mɪt dem va:gən vʁɔn ʔɛmdən naç bʁe:mən ʔnt tsuʁɪkʰ
Moderately reduced	ma:ʁn feətʰ mɪm va:gɪ vɪ ɛmdɔn naʰ bʁe:mɪn ʔn tsʁɪç
Strongly reduced	mā:m fəb mɪm vāɪɪ vɪ ɛmɪn naʰ bʁē:m ʔn tsʁɪç

Table 1 provides a summary of the three segmental reduction levels in the form of narrow phonetic transcriptions. The 9 female speakers practiced the reduction levels based on these transcriptions for several hours and then produced the statement with each reduction level several times in a row in a sound-treated recording booth of the Dept. of Linguistics at Kiel University. Recordings were made digitally at 44.1 kHz sampling rate and 24-bit quantization.

From the several tokens that each speaker produced for each reduction level, three trained phoneticians (including the author) selected, on an auditory basis, that realization that agreed precisely and most clearly with the corresponding phonetic transcription in Table 1. In this way, 9 statements (one per speaker) were selected for each of the three reduction levels, i.e. 27 statements in total.

The 27 statements were further edited with the PSOLA resynthesis of PRAAT [36] in order to create three prosodic reduction levels for each statement, thus increasing the total number of statement stimuli to 81. In parallel to the segmental domain, the three levels in the prosodic domain were also (1) *unreduced*, (2) *moderately reduced*, and (3) *strongly reduced*.

The unreduced prosody showed clear differences between stressed and unstressed syllables as well as clearly pronounced pitch-accent peaks on all content words, with H* on all pre-nuclear accents and H+L* on the nuclear accent [37]. Pitch accents were realized non-emphatically, i.e. with accent-level "2" [37], and embedded in an overall downstep pattern. The moderate reduction level left the intonation pattern intact but

largely leveled out the statement's rhythmical structure by removing the increase in duration from the stressed syllables. The strong prosodic reduction level was additionally characterized by largely flattened pitch-accent movements. Figure 1 illustrates the differences created between unreduced and strongly reduced prosodic patterns.

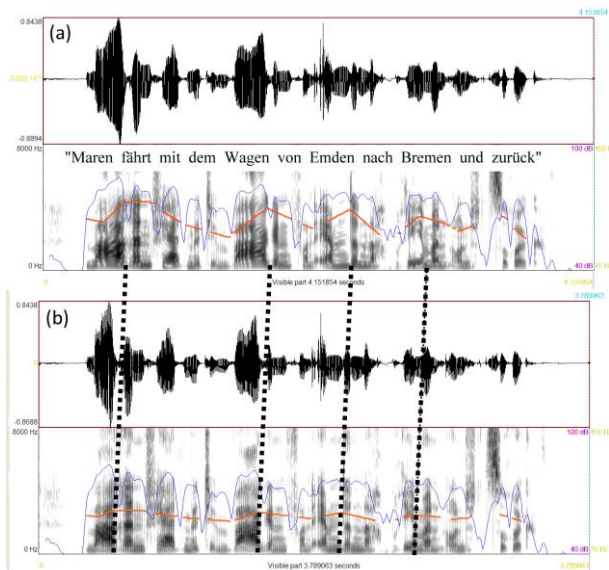


Figure 1: *Prosodic manipulation in the stimuli. Unreduced (a) and strongly reduced (b) condition of the female speaker JBE. Dotted lines indicate the shortening of lexically stressed syllables. The orange curve shows the F0 contour.*

2.3. Speaker Attributes

The stimuli were judged by the participants against 13 speaker attributes. The attributes are partly related to the Big Five (BF) personality traits [4,30], but were additionally selected such that the results would be most insightful for the applied fields of persuasive technology, social robotics, and leadership or charisma training. Correspondingly, while the BF seem to be tailored to self-assessment or traits that are primarily relevant for oneself and basically independent of one's speech, the attributes of the present study focused on traits that are relevant for (i.e. can affect) others and associated with a person's speech. Moreover, unlike the BF, the present attributes covered all three major aspects of a human being: its physical characteristics, its social characteristics, and its cognitive characteristics.

- Physical attributes: athletic, intoxicated, trained voice, clumsy.
- Social attributes (related to the "O" and "A" dimensions of the BF): sociable, educated, vain, composed.
- Cognitive attributes (related to aspects of "C", "E", "N" of the BF): stressed, tired, optimistic, sincere, scatty.

2.4. Procedure

The most important fact about the procedure is that the 3x3 segmental and prosodic reduction conditions were created for the statement renderings of each of the 9 different female speakers. This allowed compiling different stimulus sets such that each set includes all 9 reduction conditions, but each of them embodied by a *different* speaker. These sets were then distributed across the 46 listeners such that each speaker was judged, across the listeners, with respect to all 9 reduction

conditions, but by a *different* listeners for each condition. In this way, we controlled the factor '*speaker voice*' and ruled out that the results could be biased by idiosyncratic features.

Each participant heard each of the 3x3 reduction conditions 13 times; each time in combination with a different speaker attribute. This resulted in a total of 117 stimuli in the experiment. The 117 stimuli were presented in individually randomized orders. The perception experiment was carried out using a PRAAT MFC script. Participants were seated individually in silent rooms and heard the stimuli through headphones at a constant, pre-adjusted loudness level while they saw a speaker attribute on the PC screen in front of them. Each participant received 5 randomly selected familiarization trials before the actual experiment began.

At the beginning of the experiment, the participants were instructed in both oral and written form that they would hear a set of different speakers, all of them saying the same sentence: "Maren fährt mit dem Wagen von Emden nach Bremen und zurück". After having listened to each sentence, their task would simply be to judge whether or not the speaker is appropriately described by the attribute they see on the screen. Judgments were to be made by clicking with the mouse cursor on either the 'yes' or the 'no' button that were placed below the attribute on the screen. A total of 4,095 valid judgments were collected across all 46 listeners. A whole experimental session took about 20 minutes, including briefing and de-briefing.

3. Results

A three-way univariate ANOVA was used to test the effects of the fixed factors Segmental Reduction (SR), Prosodic Reduction (PR), and Attribute (AT) on the overall frequency of 'yes' responses ($n_{PR}/n_{SR}=1,365$ per factor level; $n_{AT}=315$ per factor level). The 9 female speakers were included as a covariate.

The ANOVA showed no significant separate main effect of Segmental Reduction. In contrast, there was a significant main effect of Prosodic Reduction, caused by a general decrease in 'yes' responses under strong prosodic reduction ($F[2,3977]=13.5$, $p<0.001$). That is, participants were more reluctant to assign any of the attributes to the speakers when their utterances were prosodically flattened. The interaction between Segmental Reduction and Prosodic Reduction was not significant. For the factor Attribute, the ANOVA yielded a significant main effect ($F[12,3977]=7.4$, $p<0.001$). It was caused by the fact that participants generally clicked more often on 'yes' for positively connoted than for negatively connoted attributes. That is, there was a tendency of listeners to avoid a potentially offensive judgment behavior.

The key findings of the experiment are represented by the significant interactions of the factor Attribute with Segmental Reduction on the one hand ($F[24,3977]=33.2$, $p<0.001$) and Prosodic Reduction on the other ($F[24,3977]=10.4$, $p<0.001$). The three-way interaction was also significant ($F[46,3977]=1.7$, $p=0.002$). That is, speech reduction in both domains clearly affected the perception of speaker attributes, but - as was expectable - in different directions and in different orders of magnitude for each attribute. The interaction effects are plotted in Figures 2(a)-(b). In order to examine the interactions in more detail, paired t-test comparisons between individual factors and factors levels were conducted separately for each attribute. Alpha-error levels were adjusted for multiple testing based on Bonferroni corrections. The three key findings with $p<0.05$ can be summarized as follows.

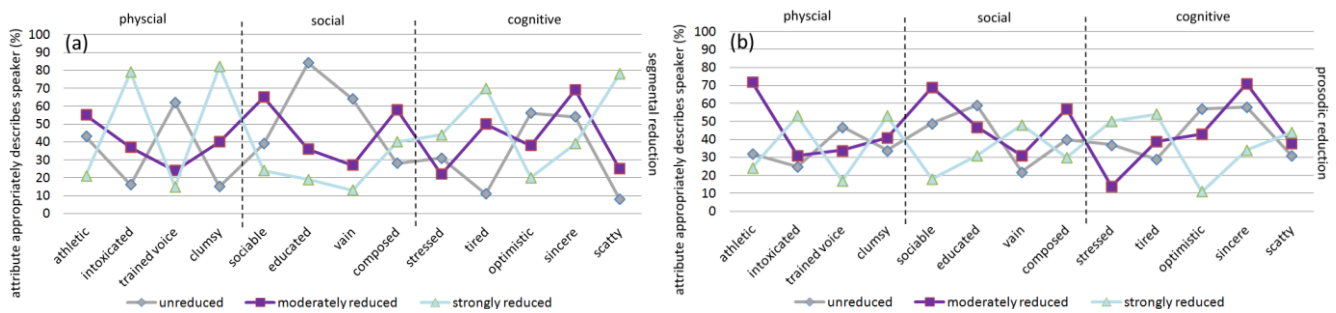


Figure 2: Relative frequencies of how often the physical, social, and cognitive speaker attributes were judged to appropriately describe the speaker. Each data point represent the 315 judgments pooled across all 46 participants.

First, in making the speakers sound more or less athletic, composed, and stressed, changes in the degree of prosodic reduction were more influential than changes in the degree of segmental reduction. In contrast, the degree of segmental reduction had larger effects than the degree of prosodic reduction on the speakers' perceived vanity, scattiness, and clumsiness.

Second, changes in segmental and prosodic reduction had parallel effects on all attributes but vanity. Moreover, the parallel effects are unidirectional for most attributes. That is, the higher the degrees of segmental and prosodic reduction were the *more* intoxicated, clumsy, tired, and scatty the speakers sounded for the listeners. Conversely, the higher the degrees of segmental and prosodic reduction were the *less* were the speakers perceived to have a trained voice and to be educated and optimistic. With respect to being athletic, sociable, composed, stressed, and sincere, changes in segmental and prosodic reduction also had parallel effects on listener judgments. However, these effects were not unidirectional. In order to sound most athletic, sociable, composed, and sincere, speakers had to show *moderate* degrees of segmental and prosodic reduction. The worst impressions in terms of these attributes were created by strong degrees of segmental and prosodic reduction. Thus, having no reduction led to intermediate results for these attributes. Similarly, sounding most stressed meant to show strong segmental and prosodic reductions. But, the least stressed sounding speakers were not those with no reduction but with *moderate* reduction levels. Again, no reduction yielded intermediate results.

Third, the perception of vanity was based on a more complex interplay of segmental and prosodic reduction. Being least vain required *strong segmental* reduction but *no prosodic* reduction. In contrast, speakers were most often judged to sound vain if they showed *no segmental* but *strong prosodic* reduction. In terms of vanity, it were the moderate reduction levels that yielded intermediate results.

4. Discussion and Conclusion

The perception experiment provides strong empirical evidence that clear speech is not mere speech, and that less clear speech is not just reduced speech either. The results revealed a complex interplay of reduction levels and perceived speaker attributes. Thus, listeners form an opinion on speakers based on their speech reduction behavior; and there is no reason to assume that the same reduction-based opinion forming does not apply in the same way also to dialogue partners in real communication situations. In this sense, the results of this experiment represent further evidence that variation in the degree of reduction is not simply determined by a tug-of-war between energy-efficient and intelligible speech signals [25].

On top of this tug-of-war, the degree of reduction is also meaningful and varied by speakers for this purpose [20].

The fact that clear speech is not mere speech is not too surprising given that production data already showed systematic links between speech reduction and individual characteristics like age, gender, and emotional state [19]. In the case of alcohol, evidence from production [38] shows a higher degree of segmental reduction for a higher degree of intoxication, and our listener judgments match well with this production evidence. The present findings also make sense in view of the effort-code concept that was put forward by [39]. Furthermore, the fact that our listeners were reluctant to assign any attribute to the speakers when their utterances were prosodically flattened is consistent with the crucial role that prosody has in conveying social and pragmatic aspects of speech [40].

The most important point in terms of the practical application of the present findings is that the link between the degree of reduction and perceived speaker attributes need not be unidirectional. It is a common assumption that well-formed speech is unreduced, and thus avoiding reductions is still considered best practice in both developing interactive machines and rhetorical training. In the latter area, for example, [41:17] states with reference to [42] that "clarity of speech and pronunciation is important for perceived charisma and influential delivery". This is not entirely wrong, of course, as our results also show that strong reduction patterns did not increase the perception of positive speaker attributes. However, a *moderate* degree of reduction clearly outperformed unreduced speech in making speakers sound more athletic, sociable, composed, and sincere. These are key attributes in social robotics and leadership/charisma training. So, future speech-synthesis developments and speech-training materials should aim at including moderate degrees of reduction in the both the segmental and the prosodic domains of speech.

Finally, the conclusions drawn from the present findings should be further substantiated and supplemented in a number of aspects. Firstly, the present stimuli were only based on female speakers; and even though there is no reason to assume that the overall results pattern will change for male speakers, it can still be expected speaker gender will make a difference, for example, due to existing gender stereotypes. This difference needs to be determined in a cross-cultural scope, and probably with an extended set of attributes. Secondly, languages other than German need to be tested, especially those languages whose reduction patterns and frequencies differ considerably from those of (West) Germanic languages like German. Thirdly, the links between reduction and speaker attributes are surely embedded in a complex interplay with semantic and situational contexts. Future studies must also shed light on this interplay. The present study only laid the first cornerstone for addressing all these questions.

5. References

- [1] Ernestus, M. (2011). Gradience and categoricity in phonological theory. In M. Van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds), *The Blackwell companion to phonology* (pp. 2115-2136). Wiley-Blackwell
- [2] L. Ellis and W.J. Hardcastle, "Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data", *Journal of Phonetics*, 30, 373-396, 2002.
- [3] F. Nolan, "The descriptive role of segments: Evidence from assimilation," In D.R. Ladd, G.J. Docherty (eds), *Papers in laboratory phonology*, Vol. 2 (pp. 261-280). Cambridge: Cambridge University Press, 1992.
- [4] R.R. McCrae and O.P. John, "An introduction to the Five-Factor Model and its applications," *J. of Personality* 60, 175-215, 1992.
- [5] D.W. Gow, "Does English coronal place assimilation create lexical ambiguity?" *Journal of Experimental Psychology: Human Perception & Performance*, 28, pp. 163-179, 2002.
- [6] O. Niebuhr and K.J. Kohler, "Perception of phonetic detail in the identification of highly reduced words," *Journal of Phonetics*, 39, pp. 319-329, 2011.
- [7] K.J. Kohler and O. Niebuhr, "On the role of articulatory prosodies in German message decoding," *Phonetica*, 68, 1-31, 2011.
- [8] P. Beddor, J.D. Harnsberger, and S. Lindeman, "Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates", *Journal of Phonetics*, 30, pp. 591-627, 2002.
- [9] S.A.J. Wood, "Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops," *J. of Phonetics* 24, 139-164, 1996.
- [10] J.J. Ohala, "Coarticulation and Phonology," *Language and Speech*, 36, pp. 155-170, 1993.
- [11] R. Pavlik, "On Defining Assimilation and Coarticulation," *Journal of Linguistics*, 60, pp. 35-56, 2009.
- [12] M. Ernestus and R. Smith, "Qualitative and quantitative aspects of phonetic variation in Dutch eigenlijk," In F. Camgemi, M. Clayards, O. Niebuhr, B. Schuppler, M. Zellers (eds), *Rethinking Reduction - Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation*. Berlin/New York: de Gruyter, 2017.
- [13] F. Cutugno, A. Origlia, and V. Schettino, "Syllable structure, automatic syllabification, and reduction phenomena," In F. Camgemi, M. Clayards, O. Niebuhr, B. Schuppler, M. Zellers (eds), *Rethinking Reduction - Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation*. Berlin/New York: de Gruyter, 2017.
- [14] K.J. Kohler, "The disappearance of words in connected speech," *ZAS Working Papers in Linguistics*, 11, 21-34, 1998.
- [15] C.G. Clopper and R. Turnbull, "Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors," In F. Camgemi, M. Clayards, O. Niebuhr, B. Schuppler, M. Zellers (eds), *Rethinking Reduction - Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation*. Berlin/New York: de Gruyter, 2017.
- [16] H. Strik, J. van Doremalen, and C. Cucchiari, "Pronunciation reduction: how it relates to speech style, gender, and age," *Proceedings 9th Interspeech Conference*, Brisbane, Australia, 2008, pp. 1477-1480.
- [17] O. Niebuhr, K. Görs, and E. Graupe, "Speech Reduction, Intensity, and F0 Shape are Cues to Turn-Taking," *Proceedings 14th Annual SigDial Meeting on Discourse and Dialogue*, Metz, France, 2013, pp. 1-9.
- [18] O. Niebuhr, "A little more ironic" - Voice quality and segmental reduction differences between sarcastic and neutral utterances," *Proceedings 7th International Conference of Speech Prosody*, Dublin, Ireland, 2014, pp. 1-5.
- [19] A. Lilliard, D. Erickson, J. De Moraes, and T. Shochi, "On the Varying Reception of Speakers Expressivity Across Gender and Cultures, and Inference in their Personalities," In S. Madureira (ed.) *Sonorities: speech, singing and reciting expressivity* (pp. 149-163). São Paulo: Universidade Católica de São Paulo, 2016.
- [20] O. Niebuhr, "Rich Reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale," In A. Botinis (ed.), *Experimental Linguistics* (pp. 11-24). Athens: University of Athens Press, 2016.
- [21] L. Plug, "From words to actions: The phonetics of eigenlijk in two communicative contexts," *Phonetica*, 62, pp. 131-145, 2005.
- [22] T. Pellegrini, A. Hämäläinen, P. Boula de Mareüil, M. Tjalve, I. Trancoso, S. Candeias, M. Sales Dias, and D. Braga, "A corpus-based study of elderly and young speakers of European Portuguese," *Proceedings 14th Interspeech Conference*, Lyon, France, 2015, pp. 852-856.
- [23] D. Byrd, "Relations of sex and dialect to reduction," *Speech Communication*, 15, pp. 39-54, 1994.
- [24] M. Kienast, A. Paeschke, and W.F. Sendlmeier, "Articulatory reduction in emotional speech," *Proceedings 6th European Conference on Speech Communication and Technology*, Budapest, Hungary, 1999, pp. 1-4.
- [25] B. Lindblom, "Explaining phonetic variation: A sketch of the H and H theory," In Hardcastle, Marchal (eds), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer, 1990.
- [26] B.J. Fogg, *Persuasive Technology: Using Computers To Change What We Think And Do*. Amsterdam: Morgan Kaufmann, 2003.
- [27] E. Strupka, O. Niebuhr, and K. Fischer, "Influence of Robot Gender and Speaker Gender on Prosodic Entrainment in HRI," *Proceedings 25th IEEE Robot and Human Interactive Communication*, New York City, USA, 2016, pp. 1-2
- [28] L.K. Bertel. PEERS: Persuasive Educational and Entertainment Robotics: A design-based research approach to social robots in teaching and learning. PhD thesis, Aalborg University, 2016.
- [29] V. Aubergé, Y. Sasa, N. Bonnefond, B. Meillon, Tim Robert, et al., "The EEE corpus: socio-affective "glue" cues in elderly-robot interactions in a Smart Home with the EmOz platform," *Proceedings 5th Emotion, Social Signals, Sentiment & Linked Open Data*, Reykjavik, Iceland, 2014, pp. 1-10.
- [30] B. Schuller et al., "A Survey on Perceived Speaker Traits: Personality, Likability, Pathology, and the First Challenge," *Computer Speech & Language* 29, 100-131, 2015.
- [31] O. Niebuhr, A. Brem, and J. Voße, "What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs' tone of voice," *Computers in Human Behavior*, 64, pp. 366-382, 2016.
- [32] S. Berger, O. Niebuhr, and B. Peters, "Winning over an audience - A perception-based analysis of prosodic features of charismatic speech," *Proc. 43rd Annual Meeting of the German Acoustical Society*, Kiel, Germany, 2017, pp. 1454-1457.
- [33] J. Antonakis, M. Fenley, and S. Lietchi, "Can Charisma Be Taught? Tests of Two Interventions", *Academy of Management Learning & Education*, 10, pp. 374-396, 2011.
- [34] G. Drosdowski. *Der Duden. Geschichte und Aufgabe eines ungewöhnlichen Buches*. Mannheim: Dudenverlag, 1996.
- [35] K.J. Kohler. *Einführung in die Phonetik des Deutschen* (2nd ed.). Berlin: Erich Schmidt, 1995.
- [36] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, 5, pp. 341-345, 2001.
- [37] S. Baumann, O. Niebuhr, and B. Schroeter, "Acoustic Cues to Perceived Prominence Levels: Evidence from German Spontaneous Speech," *Proceedings 8th International Conference of Speech Prosody*, Boston, USA, 2016, pp. 711-715.
- [38] C. Gussenhoven, "Intonation and interpretation: Phonetics and Phonology," *Proceedings 1st International Conference of Speech Prosody*, Aix-en-Provence, France, 2002, pp. 47-57.
- [39] H.J. Künzel, A. Braun, and U. Eysholdt. *Einfluß v. Alkohol auf Sprache und Stimme*. Heidelberg: Kriminalistik-Verlag, 1992.
- [40] A. Mokhtari & N. Campbell, "Speaking Style Variation and Speaker Personality," *Proceedings 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, pp. 1-4
- [41] L. Lourekai. *Charismatic communication style in knowledge-intensive organizations*. MA thesis, School of Business, Aalto University, Finland, 2015.
- [42] J.M. Kouzes and B.Z. Posner. *The leadership challenge*. New York; Wiley & Sons, 2003.