Politeness, culture, and speaking task – paralinguistic prosodic behavior of speakers from Austria and Germany

Sven Grawunder[1], *Marianne Oertel*[2], *Cordula Schwarze*[3]

[1]Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]Department of Speech Science, Martin Luther University Halle-Wittenberg, Halle, Germany
[3]Department of German Studies, University of Innsbruck, Austria
grawunder@eva.mpg.de

Abstract

This paper tests previous findings for polite speech of low pitch, low intensity, higher number of hesitation markers and filled pauses against those parameters in a different socio-cultural background. Two similar groups of (19+13) participants, from Austria and from Germany, were recorded. The adopted experimental approach used 16 tasks aiming at different speech acts in situations that evoke either polite or informal speech. The analyzed acoustic and electroglottographic signals reveal main effects for lower pitch, lower intensity and HNR only for the German group. Open quotient values differ only for female speakers. In both groups significantly lower word rate and lower speaking rate as well as higher rates of filled pauses and hesitation markers are found in formal (polite) conditions. However individual speakers can show indifferent or opposing behavior for a given parameter with compensatory utilisation of other parameters in order to express politeness (formality).
Index Terms: polite speech, voice quality, pauses, hesitations

1. Introduction

While looking at expressions on the sociopragmatic level of speech [1] a number of phonetic and other paralinguistic differences between polite vs. informal speech registers have been found for situations talking either to a senior person or to a peer. We adapt the view that the concept of polite speech covers here typically a transmission of a social message of deference, submission and distance, where politeness (polite behavior) is considered to be a vehicle to ensure stability of social hierarchies. In the model of Brown & Levinson [2] this is categorized as negative politeness strategy. Such speaking style is different from the speech of, e.g., shop assistants or other service personel, that seeks an exaggeration of displayed interest, approval, or sympathy with the interlocutor, which is identified as positive politeness [2]. And in fact, Ito [3] and Ofuka [4] found for Japanese female speakers that high pitch register and higher degrees of breathiness correlate with higher degrees of perceived politeness. Higher pitch range in connection with perceived politeness are reported for British English and Dutch by Chen et al. [5] but also for a number of Spanish Varieties in Latin America (cf.[6]). However, for Korean females a pitch lowering was found [7] in polite speech, which was confirmed in a independent production study and also found for male speakers [8]. In a recent study it was shown that English and Korean listeners can detect the intended politeness purely based on prosodic expressions (in Korean), even if only presented with minimal material like a sentence [9]. Accuracy would increase for native and non-native listeners as they become familiarized with the voices. Similarly a reduction of pitch range in Catalan [6] has been found to prompt a perception of politeness. These findings do not meet the assumptions of high pitch as known from the predictions by Brown & Levinson [2] or the frequency code hyphothesis of Ohala [10, 11], but suggest a more differentiated hypothesis with regard to the phonetic profile of politeness. Additionally it brings into play other biologically motivated sound symbolisms in human and non-human vocal behavior, namely the effort code hypothesis by Gussenhoven [12], since here other voice quality parameters apart from pitch are used. Hence this paper seeks to contribute more rigorously to the question of cross-linguistic validity of such findings and seeks to replicate the previous study [8] with a different cultural and linguistic background. In order to address also the cultural pragmatic aspect, we conducted the same experiment in two linguistic expressions of cultures (Austria and Germany) [13, 14] of the same language (German). If a more universal usage of a "frequency code" holds we would – given the results for Korean [8, 9, 15] or Catalan [6] – expect similarily for polite speech: lower F0, lower F0 range, lower intensity, lower intensity range, more tensed VQ with parameters indicating this (i.e. acoustically, lower harmonics-to-noise ratio (HNR), or electroglottographically, lower open quotient). Further we would expect for polite speech more hesitation phenomena (onset hesitation, lengthening, hesitation 'markers') and more signals of interruption (audible breath intakes etc., cf.[8]).

2. Data and Methods

2.1. Participants and Procedure

The current data set comprises recordings of 13 speakers (11 female/2 male) from East Central German and of 18 speakers (8 female/10 male) from Western Austria. Recording sessions were carried out in the phonetics lab of the MPI EVA Leipzig and in the sound attentuated room of the Media Center of University of Innsbruck. For best replication of the previous study, we were aiming at students of the humanities as participants which were not freshmen any more but still in the age of 20 to 30. The latter ensured that our participants already had commuciation experience in hierarchies and asymmetrical relations, but were not too advanced in their carrier. To ensure consistency of procedure, participants were first ask to read aloud a written text in order to get acquainted with the setup. After clarifying the general procedure the participants were prompted with different situations via a screen. The 16 situations described demanded either a mailbox task, where a participant would be asked to leave a message for her friend or her boss on a voice

10.21437/SpeechProsody.2014-20

mail box. Or the task was a discourse completion task (DCT), where a participant would be asked to start a conversation, like *'Remind your professor to upload the power point files after the lecture'*. Linguistically the scenario descriptions were culturally adapted for the local areas in Austria and Germany so that infrequently used phrases and non-typical names were avoided.

2.2. Recordings and Measures

The recordings of the acoustic (headset microphones AKG C420 III and audix HT5) and electroglottographic signals (Glottal enterprise EG2) were accomplished by means of a multitrack recorder sound device 788T in a PCM format of 44.1kHz 16bit. *Praat* [16] was used as annotation tool and for acoustic and EGG measures. Manual annotation and labeling was carried out in order to assess word rate, occurence of hesitation markers as well as audible and silent breath intakes. Syllable estimation [17], f0, f0 variance, f0 range, intensity (RMS), intensity range, HNR as well as noise-to-harmonics ratio (NHR), as well as jitter (local, RAP) and shimmer (local, APQ3) [16, Manual] measures were based on the periodic parts of the microphone signal. The amplitude difference H1-H2 had not been chosen, since especially for analyzing running speech these have been lately critized to be heavily vowel dependent [18]. The EGG-signal served for the open quotient, i.e. duration of open phase by total period. These quotients were based in part on the DEGG-method where the first derivative is used for estimation of the instants of closing and opening (cf. [19]).

2.3. Notes on Statistics

We used *R* [20] with the packages *lme4* [21] and *glmmADB* [22] to perform a linear mixed effects analysis of the relationship between the individual parameters and the supposed attitude (polite vs. informal). In a first step we set attitude and gender as fixed effects into the model, in a second step we added group (Austrian, German) to the fixed effects. As random effects, we had random intercepts for speakers and tasks, as well as random slopes for the effect of attitude by speaker and by task. Visual inspection of residual plots served as check for deviations from homoscedasticity or normality, including overdispersion by particular random effects (task, speaker). P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Likewise models with interaction (e.g. attitude*gender) were tested against such models without interaction (e.g. attitude+gender). For the count data (pauses etc.) we were using with negative binomial and zero.inflation=TRUE.

3. Results

3.1. Polite vs. Informal: Overall

3.1.1. General Observations

All participants reacted eagerly to solve the experimental tasks quickly and freely. Specifically the mail box task proved to be appropriate, especially in initiating the experiment and gaining near real-life utterances for most of the time. The formal vs. informal distinction was realized by all participants. It was not only that participants would choose the correct grammatical forms (e.g. du 2SG 'you; thou' vs. Sie 'you' 3SG) for addressing the imaginary interlocutor, but they would also use more passive or subjunctive constructions. With regard to pronunciation we almost always observed a shift away from close-to-standard patterns into to more areal or colloquial patterns in informal situations. This was more clearly obvious for the Austrian group, since the (spoken) standard for all German speaking countries is derived from more Northern German pronunciation habits, hence there is a greater phonetic distance between the two varieties and therefore the two situational styles/registers.

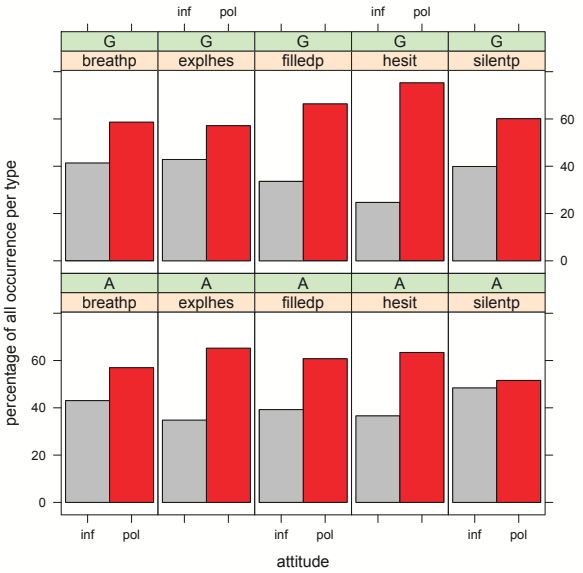

Figure 1: *occurrence of breath, silent and filled pauses as well as explicit and other (lengthening) hesitation markings, grouped by country (*Austria*; Germany)*

3.1.2. Hesitation Phenomena and Rates

One of the most striking phenomena is the occurence of hesitation phenomena throughout the corpus. In order to group the different appearances of these phenomena in a meaningful way we defined five types. First, we separated pauses into *silent pauses* and (audible) *breath pauses* (in and out). Then we would treate conventionalized hesitation markers – also known as non-lexical conversational sounds (cf. [23]) or fillers – like "ä","äh","ähm" ('hm, mh , em' etc.) as *filled pauses*. In contrary such coventionalized explicit phrases that contain lexical items like "ich denk", "glaub ich" ('I guess, I think') or conjunctions like "u:::nd", "dann:: ", "weil::" ('and, well, because' etc.) – functioning as discourse markers – were grouped as *explicit hesitations*. Although one can observe all kinds of combinations of the conventionalized markers, phrases and conjuctions plus a phonetic lengthening (like "u:::nd", "dann:: ", "weil::" or "und ähm", "weil ähm", "dass ähm") we nonetheless treated these lengthening behavior as one *hesitation* category. Silent pauses would also comprise instances of breath holding, i.e. a pause with oral or glottal closure, with a more or less sudden release. These cases were also treated under the same category, *hesitation*.

Overall we find higher counts and percentages for all types of hesitation marking and pauses in the formal (polite) condition. Breathing pauses (inhales) show a significant main effect (attitude=polite, Est. +0.363 ±2.99 SE, p=0.0028) with no (gender) interactions. However one needs also to take into account that the answers in the polite condition have a longer duration. The relation of effective duration and total duration of a task reflects the role of pauses per task similar to the rela-

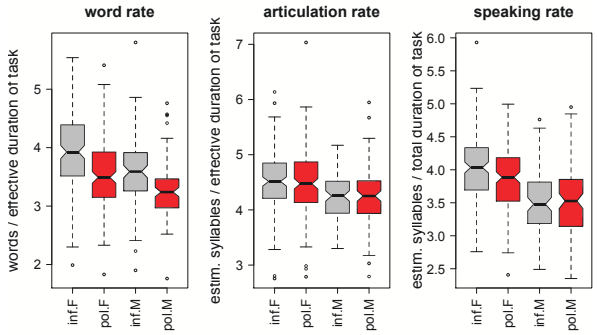

Figure 2: *Overall rates for all female (F) and male (M) speakers in polite (pol) and informal (inf) condition*

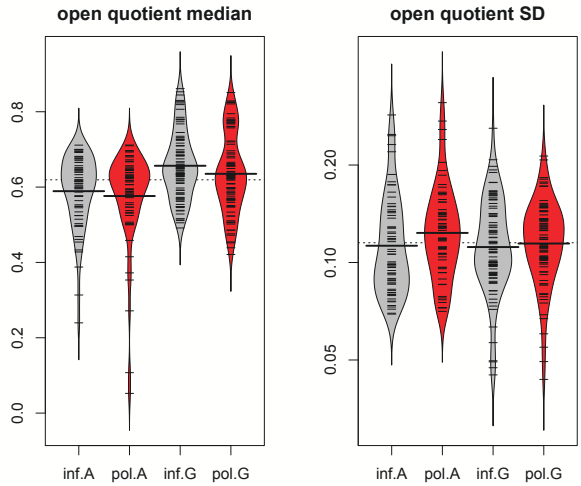

Figure 3: *open quotient measures for female participants from Austria (A; N=8) and Germany (G; N=11)*

tion of speaking rate vs. articulation rate (-0.025s ± 0.01154, p=0.037). Closely related are the results of word rates (p=0.03), since here a higher number of pauses corroborates the lower rate of words (per task duration). Moreover the lower 'wordiness' may even be more pronounced since here all non-lexical markers have not been excluded yet.

3.1.3. Pitch, Intensity and Voice Quality Measures

Average f0 values display only a minimal overall main effect of -0.41st and ±0.217, i.e. lower f0 in the polite condition. Only half which comes out significant for interaction (attitude*gender) for the German subgroup (-0.44st ±0.273, p=0.012). Further only mean intensity differences reach significance in interaction (-1.24dB ±0.460, p=0.02746). The only acoustic voice quality parameter showing significant effects reflects the variance of HNR within a task: overall HNR SD is lower for polite speech with interaction effect (0.28dB ±0.116, p=0.019), in the Austrian (0.17dB ±0.109, p=0.023) and the German group (here only as main effect, 0.14dB ±0.098, p=0.015). This points towards a more non-homogenous use of voice quality on the breathiness-tensedness dimension. Taking into account correction for multiple testing we note that this result is corroborated by the results for open quotients (OQ; Fig. 3), where we see lower values in polite situations (mean OQ: 0.013 ±0.0114, p=0.0364; OQ SD: 0.0105 ±0.0041, p=0.043; here only reported for interaction with country since only female speakers were compared). These findings suggests a higher tensedness (or lower breathiness) for polite speech and thus far parallel the results for Korean [8].

3.2. Group specific behavior

3.2.1. Overall gender specific behavior

Gender specific main effects are found for parameters that are expected and well known to differ between genders (sex). These concern mainly pitch and voice quality (jitter, shimmer), where f0 range seems to be strikingly higher for male speakers in both conditions. As often observed, we find higher articulation rates for female speakers (Fig. 2; with interaction, 0.31words/s ±0.160, p=0.00014). More interesting in this regard is that word rates show a signficant effect with interaction for lower rates in polite speech with male speakers (0.013words/s ±0.080, p=0.039).

3.2.2. German in Austria vs. German in Germany

Apart from a general trend for open quotients, which appear to be very small but significantly higher for female speakers from Germany (mean OQ:-0.0024 ±0.01806, p=0.039) than for female speakers from Austria, meaning that there comes into play a difference between these groups, perhaps pointing towards a general use of breathiness in the speaker population from Germany. Other measures where we find group (country) effects but also main effects for attitude are mean intensity values (-6.46dB ±1.542, p=0.00011) and HNR SD values (0.79dB ±0.3106, p=0.0053). With regard to the nature of these measures such differences may be easily explained by the influence of conditions (e.g. mic or booth) in the two recording sites.

3.3. Task specific behavior

The individual task categories (eight scenarios) that we were testing for phonetic aspects comprise a number of different speech acts, i.e. types of messages that have to be conveyed by the paticipant. These required urgently requesting an action (task 11, 12), requesting something critical (task 2,6), requesting something essential (task 1), apologizing for being late, making a compliment (task 5), and correcting a (minor/major) mistake (task 3,4). Indeed we observe, more often especially for task 5 (*compliment*) but also for task 6 (*minor request*), behavior that deviates from that of other tasks. For mean HNR measures in task 5, showing the lowest values in the dataset, a three way interaction of attitude, gender and task becomes marginally significant (-1.42dB ±0.737, p=0.0545). Moreover, mean HNR measures show for task 2 and 12, lower values for males and females whereas the other tasks display the contrary trend, e.g. males in task 3 (1.48dB ±0.521, p=0.0045). Strikingly the two groups behave almost symmetrically (s. Fig. 4) and we are not yet able to discard a particular task as behaving consistently different.

3.4. Speaker specific behavior

While looking at strategies of individual speakers we report in Table 1 individual slopes for a selection of parameters that show partially an opposing behavior (cf. [24]). For example

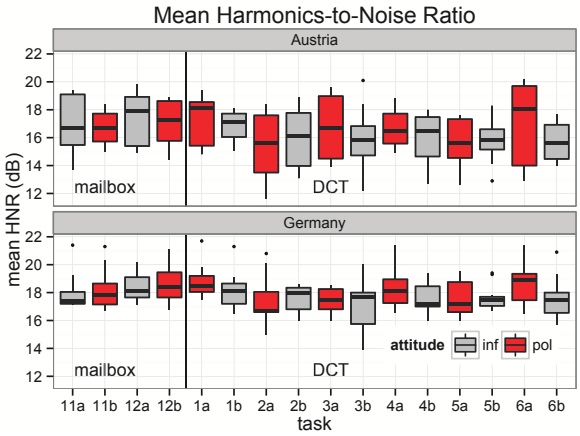

Figure 4: *taskwise HNR behavior (females only)*

speaker GM01 illustrates the complementary use of high intensity range difference but countertrend f0 range. At the lower end we find speaker AM02 with the opposite behavior. Similarily HNR seems to be more strongly employed (e.g. by GM01). By contrast, all speakers exhibit the same trend for the pause quotient (effective task duration by total task duration).

4. Discussion

On the one hand the current results for speakers of two German varieties show tendencies for the use of hesitations and pauses as well as for the use of pitch, pitch range, intensity, intensity range, and HNR, that match those results for Korean in [8]. On the other hand, the two speech registers appear to be less distinct, given the lower number of significant differences, which can be assumed to mirror the fact that for German politeness or formality register is linguistically but also culturally less entrenched. Certainly we must stay aware of the reality of the experimental nature of the investigated speaking behavior. The general assumption here is that participants will simulate from the same source of acquired culturally transmitted and evolutionary conditioned behavioral pattern as they would in a non-experimental situation of a similar situation 'in situ'. The latter would then supposedly evoke higher arousal levels and lead to stronger difference between the attitudes (formal/informal). We cannot exclude at this moment that individual speakers may also fail to express deference and submission widely in favour of other expressions, such as friendliness or fear, while still being formal. This could in case of friendliness be interpreted as pursuing a positive politeness strategy, or in case of fear, as an exaggeration of the urgency message. Both may remain disrespectful and therefore potentially impolite if not addresssed to peers or if failing formality even with addressing a peer – a distinction that remains to be tested yet.

5. Conclusions

Since languages vary in their entrenchment of pragmatically relevant registers (such as different degrees of politeness) into the grammar, moreover the paralinguistic expression may vary between speech communities. Small scale samples, as ours, may not give enough power to estimate the actual group differences that interact with gender specific effects. Within communities we observe speakers using the multidimensional space of possible parameters (pitch, loudness, voice quality, fluency; cf.[25]) in its own particular way but also in the degree of expression. Already previous results for Korean [8] contradict the findings for Japanese and English [3, 4, 5, 7] where politeness was associated with higher pitch (and more breathiness). Although with our German speaking participants we find more non-homogenous behavior of subjects, our findings are still very similar, since we find clear trends for (less breathy) voice quality and higher rates of pausing/hesitations in polite speech. But we can also confirm the previous observations of Shin [7], who found no pitch differences for German. Given that we were using here still a parsimonious approach, using 'holistic' rather than detailed measures – an in-depth analysis of the actual moments of expression for pitch [6], intensity or breathiness/tensedness will certainly gain more insigths. In order to achieve ecological validity we will also turn these production data into a perception experiment.

Table 1: *Random slopes per speaker from polite condition sorted by f0 range values; pausequot (effective task duration / total task duration)*

speaker	f0 range	Int range	pause quotient	mean HNR
GM01	-0.08	2.65	-0.024	0.65
AM03	-0.01	1.18	-0.026	0.33
AM06	0.02	0.69	-0.021	0.36
AF09	0.76	1.92	-0.025	0.48
GF11	1.12	0.17	-0.026	0.49
AM08	1.18	0.40	-0.027	0.04
AF06	1.30	1.34	-0.023	0.19
AM07	1.37	0.81	-0.026	0.40
AF08	1.51	1.98	-0.023	0.49
GF05	1.62	2.08	-0.026	0.53
AM05	1.89	1.05	-0.030	0.31
GF03	1.91	1.89	-0.026	0.58
GF08	1.93	1.08	-0.027	0.30
GF04	2.04	2.25	-0.024	0.23
AF04	2.04	0.26	-0.027	-0.01
AF01	2.04	1.08	-0.027	-0.09
AF03	2.05	-0.31	-0.026	0.30
GF09	2.14	0.28	-0.023	0.27
AF02	2.17	1.14	-0.023	0.52
GF10	2.31	0.13	-0.027	0.22
AF07	2.55	0.68	-0.025	0.08
GF06	2.55	1.06	-0.023	0.53
AM10	2.67	1.34	-0.019	0.30
GF07	2.69	1.44	-0.027	0.38
GF02	3.03	1.29	-0.025	0.67
AM02	3.21	-0.08	-0.027	0.17
GF01	3.23	0.27	-0.026	0.33
GM02	3.49	1.14	-0.026	0.58
AM01	3.60	1.69	-0.027	0.34
AM09	3.61	0.42	-0.023	0.38
AM04	3.67	1.36	-0.025	0.24

6. Acknowledgements

We would like to thank all our participants and express our gratitude to Leonardo Lancia (MPI EVA), to Anton Tremetzberger (Uni Innsbruck) for assistance during the recording sessions, to Yvonne Kathrein for helping with the 'Austrification' of our stimuli and to Bodo Winter (UC Merced) for his technical advice. Finally we would like to thank the MPG, specifically Prof. Bernard Comrie for his support.

7. References

[1] M. Tatham and K. Morton, *Expression in speech: Analysis and synthesis*. Oxford: Oxford University Press, 2004.

[2] P. Brown and S. Levinson, *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press, 1987.

[3] M. Ito, "Politeness and voice quality–the alternative method to measure aspiration noise," in *Proceedings of the Second International Conference on Speech Prosody, Nara, Japan*, 2004, pp. 213–216.

[4] E. Ofuka, J. D. McKeown, M. G. Waterman, and P. J. Roach, "Prosodic cues for rated politeness in Japanese speech," *Speech Communication*, vol. 32, no. 3, pp. 199–217, 2000.

[5] A. Chen, C. Gussenhoven, and T. Rietveld, "Language-specificity in the perception of paralinguistic intonational meaning," *Language and Speech*, vol. 47, no. 4, pp. 311–349, 2004.

[6] M. Nadeu and P. Prieto, "Pitch range, gestural information, and perceived politeness in Catalan," *Journal of Pragmatics*, vol. 43, no. 3, pp. 841–854, 2011.

[7] S. Shin, "Grammaticalization of politeness: A contrastive study of german, english and korean," PhD Thesis, University of California, Berkeley, 2005.

[8] B. Winter and S. Grawunder, "The phonetic profile of Korean formal and informal speech registers," *Journal of Phonetics*, vol. 40, no. 6, pp. 808–815, 2012.

[9] B. Winter, L. Brown, K. Idemaru, and S. Grawunder, "Perceiving politeness from speech acoustics alone: A cross-linguistic study on Korean and English," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 4072–4072, 2013. [Online]. Available: http://scitation.aip.org/content/asa/journal/jasa/134/5/10.1121/1.4830871

[10] J. J. Ohala, "An ethological perspective on common cross-language utilization of f0 of voice," *Phonetica*, vol. 41, pp. 1–16, 1984.

[11] ——, "The frequency code underlies the sound symbolic use of voice pitch," in *Sound symbolism*, L. Hinton, J. Nichols, and J. J. Ohala, Eds. Cambridge: Cambridge University Press, 1994, pp. 325–347.

[12] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[13] S. Haumann, U. Koch, and K. Sornig, "Politeness in Austria: Politeness and Impoliteness," in *Politeness in Europe*, L. Hickey and M. Stewart, Eds. Multilingual Matters Clevedon, 2005, pp. 82–99.

[14] J. House, "Politeness in Germany: politeness in Germany," in *Politeness in Europe*, L. Hickey and M. Stewart, Eds. Multilingual Matters Clevedon, 2005, pp. 13–28.

[15] L. Brown, B. Winter, K. Idemaru, and S. Grawunder, "Phonetics and politeness: Perceiving Korean Honorific and Non-Honorific Speech trough Phonetic Cues," *Journal of Pragmatics*, 2014, accepted.

[16] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*, Version 5.3.10, retrieved 12 March 2012 from http://www.praat.org/, 2012.

[17] N. H. de Jong and T. Wempe, "Automatic measurement of speech rate in spoken dutch," *ACLC Working Papers*, vol. 2, no. 2, pp. 49–58, 2007.

[18] A. P. Simpson, "The first and second harmonics should not be used to measure breathiness in male and female voices." *Journal of Phonetics*, vol. 40, pp. 477–490, 2012.

[19] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo, "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–32, 2004.

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org/

[21] D. Bates, M. Maechler, and B. Bolker, *lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0*, http://CRAN.R-project.org/package=lme4, 2012.

[22] D. A. Fournier, H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert, "Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models," *Optimization Methods and Software*, vol. 27, no. 2, pp. 233–249, 2012.

[23] N. Ward, "Non-lexical conversational sounds in american english," *Pragmatics & Cognition*, vol. 14, no. 1, pp. 129–182, 2006.

[24] K. Drager and J. Hay, "Exploiting random intercepts: Two case studies in sociophonetics," *Language Variation and Change*, vol. 24, no. 1, p. 59, 2012.

[25] N. Campbell and P. Mokhtari, "Voice quality: the 4th prosodic dimension," *15 th International Congress of Phonetic Sciences*, pp. 2417–2420, 2003.