Acoustic correlates of politeness: prosodic and voice quality measures in polite and informal speech of Korean and German speakers

Sven Grawunder[1], *Bodo Winter*[2]

[1]Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]Department of Linguistics, University of Hawai'i at Manoa, USA

grawunder@eva.mpg.de, winterbo@hawaii.edu

Abstract

This paper investigates phonetic and prosodic features of polite versus informal speech. Two different communicative tasks were performed by speakers of two unrelated and culturally distant languages, Korean and German. We found that the polite speech of Korean speakers can be characterized by an increase of filled pauses and extralinguistic markers, a higher degree of breathiness, as well as lower measures of average fundamental frequency, intensity, period perturbation and amplitude perturbation. We also found that German polite speech shares similar tendencies. The parameters studied seem to contradict previously claimed universals about polite speech.

Index Terms: politeness, voice quality, Korean, German, gender

1. Introduction

Fundamental frequency (f0) has been hypothesized to serve as a key marker for a number of paralinguistic and extra-linguistic functions. Approaching the role of fundamental frequency from an ethological perspective, Ohala [8, 4] quotes Bolinger [3] who notes that "it seems safe to conclude that such 'social' messages as deference, politeness, submission, lack of confidence, are signaled by high and/or rising f0 whereas assertiveness, authority, aggression, confidence, threat, are conveyed by low and/or falling f0". Brown & Levinson posit two types of politeness ("positive politeness" or politeness-as-sympathy; and "negative politeness" or politeness-as-deference) and predict that across different cultures, positive politeness will be associated with creakiness and negative politeness with relatively high f0 [4, 267-268]. They also predict that "the reversal of these associations will not occur in any culture". The claims made by Bolinger and Brown & Levinson were made before cross-cultural comparisons of phonetic parameters have been undertaken. This highlights the role of further exploratory work into what exactly constitutes prosodic and phonetic characteristics of politeness.

Campbell & Mokhtari [5] demonstrated that next to f0, voice quality measures play a crucial role in social interaction. Therefore, we investigate a variety of voice quality parameters such as shimmer, jitter, NAQ and H1-H2. Since Ito [6] demonstrated that in production as well as in perception, breathiness is associated with politeness in Japanese, we will pay particular attention to acoustic correlates of breathiness. We decided to focus on Korean because of its highly elaborate system of grammaticalized politeness markers (so-called honorifics). Also, Korean has two clearly delineated cultural categories, *contaymal* (polite/formal speech) and *panmal* ('half-speech' or informal speech), which allow us to operationalize politeness for the purpose of this study. Since these categories are 'culturally entrenched' and reinforced through education in Korea, we make it less difficult for our participants to switch politeness registers in a laboratory situation. We contrast the speaking behavior of Koreans with those of German speakers in order to find similarities which might point to universal tendencies of how politeness is phonetically implemented.

2. Method

2.1. Note Task and Discourse Completion Task

Both groups of speakers (Korean & German) were presented with two types of task. In the Mailbox Task (MBT), the participants were asked to leave a pre-specified message on a voice mail box for either a friend or a professor (both were fictitious). In the Discourse Completion Task (DCT), participants were asked to role-play the beginning of a dialogue with several different kinds of speech acts (request, apology, warning, correction, appraisal). The task descriptions and context scenarios are largely comparable between the two groups (Korean/Germans), however, some slight cultural adjustments had to be made. In total, each participant was presented with 2 dyads of informal/polite MBT contexts and with 5 dyads of DCT contexts. All recordings were done via a head-set microphone AKG C420 (linear characteristic) with 48kHz /16bit sampling in a sound booth. The subjects were presented with a description of the task in form of a short text, and then a picture of the imagined interlocutor appeared on a screen in the booth.

2.2. Participant Speakers

16 Korean students (9 female/7 male; music/humanities) were recruited in Germany. The majority (13) of Korean speakers came from the Seoul metropolitan area and reported to be speakers of the Seoul dialect. In addition, we recruited 9 female students (humanities) who were native speakers of German (5 from Halle-Leipzig, 2 from Cologne, 1 from Hamburg and 1 from Osnabrück). All participants of both groups were between 20 and 30 years old.

2.3. Acoustic Parameters of Voice Quality

Voiced parts were extracted and concatenated to a single sound file. From this file, average f0 (with the Praat autocorrelation pitch tracking; standard settings adjusted for gender) and intensity were measured over the total duration of each task. We calculated f0 and intensity ranges with the difference between the 0.9th and 0.1th quantile of each task. Variability of f0 was calculated with the standard deviations (SD) of each task.

For the assessment of the overall amplitude and period variation within each task, we chose several relative measures of jitter and shimmer which cover different scopes of the fundamental frequency: Local jitter and shimmer cover 2 adjacent cycles, RAP and APQ3 cover 3, PPQ5 covers 5, and APQ11 covers 11 cycles [2, Manual]. In order to assess the breathiness of our participants' responses, we calculated Harmonics-to-Noise Ratio (HNR)[2, Manual] and the amplitude difference of the first and second harmonic (H1-H2) [7]. All voice quality measurements were conducted on the voiced-only sound files mentioned above.

The NAQ (Normalized Amplitude Quotient) introduced by Alku [1] comprises the "ratio between the amplitude of the ac flow and the negative peak amplitude of the flow derivative normalized by the period length of the fundamental" [1, 701]. NAQ is a relatively robust indicator of voice qualities which span a continuum from pressed voices to breathy voices. It has a good sensitivity for tensed voices, is relatively independent from f0 and has been found to vary with respect to such social factors as the status and the familiarity of a speaker's interlocutor [5]. For the NAQ analysis, we only considered fillers with an [O]-like vowel quality. In each condition (polite / informal), we chose one task of which 6 to 8 vowel sequences of 100 to 400ms were selected. The inverse filtering for each of the recombined sequences was carried out by means of WAVEVIEW PRO 2.3.1 developed by Martin Rothenberg (*Glottal Enterprises*), with 50Hz low pass filtering and a specific but fixed formant adjustment. The after-processing, including NAQ derivation, detection of extrema and measurement was done in PRAAT [2].

2.4. Non-Acoustic Phonetic Parameters & Statistics Report

Whereas we provide analyses of the acoustic parameters for both groups, only the Korean group was analyzed with respect to non-acoustic parameters. For this group, we look at three measures of speech rate – syllables per second (syllable rate), syllables per second excluding pause times (articulation rate) and words per second (word rate), as well as rates of pauses and fillers for the DCT. A syllable was defined as a single character of the Korean *Hangul* transcript. For the pause analyses, quasi-zero amplitude sequences of a minimum of 200ms were interpreted as silent pauses. We also counted the number of breath intakes and hesitation markers such as *hm*.

We performed repeated-measures ANOVAs with attitude as a within-subjects effect and gender as a between-subjects effect for each measured parameter. We tested attitude as main effect against the interaction with subject /task, as well as gender effects, and subject against the error. We report two F-ratios, F1 for participants [subjects] and F2 for task [item].

3. Results

3.1. General Observations

Speakers were able to adapt to both tasks. The Korean participants used the grammatical markers of *contaymal* and *panmal* appropriately. The phonetic and prosodic differences between these two categories are audible. The overall auditory impression of the polite speech register is that of a more monotonous speaking style with a decrease of loudness, a sometimes more tensed voice quality and 'streamlined' pitch movements. Furthermore, there is a salient increase of fillers and laterally produced breathing sounds in the polite speech register.

The phonetic politeness distinctions of the German speakers are less apparent even though the phonetic measurements point into similar directions as the phonetic measurements of the Korean speakers, especially with regard to decrease of loudness and monotony of pitch.

3.2. Fundamental Frequency & Intensity

For all averaged measures of f0, male and female subjects tend to show higher values for informal speech (*panmal*) than for polite speech (*contaymal*) in a range of about 1-3 semitones. 15 of 16 subjects showed higher median f0 and 14 showed higher mean f0 values. In semitone scaled measures we gained a highly significant difference between the two politeness registers (F1(1,15)=29.13, p<0.001; F2(1,6)=14.83, p=0.008) with a natural gender effect but no interaction between gender and politeness register. In the polite speech register, we also observed a clear preference for a decrease in f0 range for female speakers (8 female) but not as clear for male speakers (only 3 males). Furthermore, for 9 of 16 speakers, the f0 SDs show higher median values in informal speech, which indicates the more 'lively' and less monotonous character of informal as opposed to polite speech. We see significant effects for attitude in period SD (F1=29.23, p=0.003; F2=2.39, p=0.043). The German speakers behave similarly to the Korean speakers – 8 of 9 speakers reveal higher median and mean f0 values in informal speech. Four speakers also show higher f0 rangers when speaking politely.

As for loudness, the intensity values of the majority of Korean speakers (13 of 16 for mean intensity) tend to be lower in polite than in informal speech (two speakers showed significant differences p<0.05). The ANOVA reveals significant differences of subject-wise analysis (F1=20.05, p1=0.001; F2=2.1, p2=0.198). Similarly, 6 of 9 German speakers show lower values in polite speech (F1(1,8)=6.16, p1=0.038; F2=(1,6)2.13, p2=0.195).

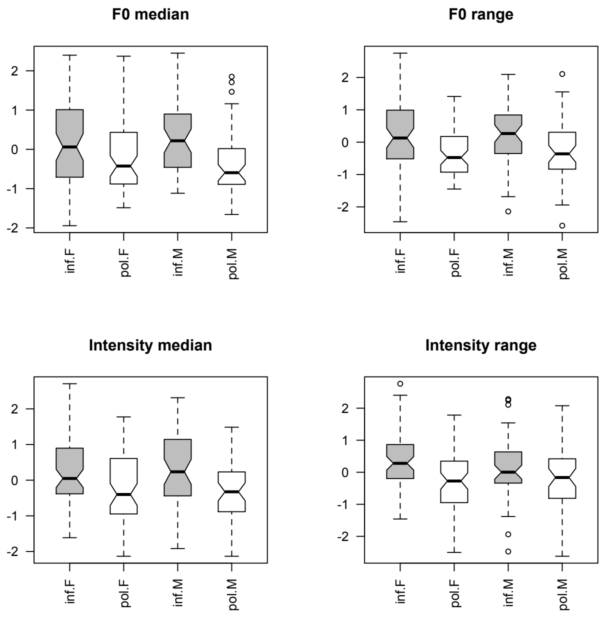

Figure 1: *Median F0, F0 range, median Intensity, and Intensity range as z-scored values grouped by gender (F =female, M=male) and mode (inf=informal, pol=polite); Korean*

These findings are consistent with Shin's findings on Korean politeness [11], but they are different from the results of studies conducted on Japanese politeness [9]. The findings also stand against the predictions made by Brown & Levinson [4, 267-268] as well as against the predictions made by Ohala [8] and Bolinger [3].

Table 1: *Summary of results of participant-wise (F1(1,16)) and task-wise (F2 (1,6)) ANOVAs for Korean speakers; with attitude (att: informal vs. polite) as main within effect, gender (gen), and their interaction (int)*

subject analysis						
Parameter	F1(gen)	**F1**(att)	F1(int)	p1(gen)	**p1**(att)	p(int)
f0 median	82.41	29.13	5.59	0.000	0.000	0.033
period SD	29.23	12.53	5.52	0.000	0.003	0.034
local jitter	33.14	11.31	2.26	0.000	0.005	0.155
jitter RAP	18.82	1.43	6.86	0.001	0.251	0.020
jitter PPQ5	24.03	4.51	0.98	0.000	0.052	0.339
local shimmer	2.59	30.77	3.24	0.130	0.000	0.093
shimmer APQ3	0.02	18.88	1.75	0.893	0.001	0.207
shimmer APQ11	14.18	24.31	0.01	0.002	0.000	0.945
HNR mn	5.13	56.48	0.69	0.040	0.000	0.420
intensity mn	2.12	20.05	3.48	0.167	0.001	0.083
H1-H2	14.91	17.58	5.2	0.002	0.001	0.039
items analysis						
Parameter	F2(gen)	**F2**(att)	F2(int)	p2(gen)	**p2**(att)	p(int)
f0 median	1526.86	14.83	0.21	0.000	0.008	0.658
period SD	337.51	2.39	3.09	0.000	0.173	0.104
local jitter	405.67	3.12	0.06	0.000	0.128	0.810
jitter RAP	1089	1.01	1.46	0.000	0.354	0.250
jitter PPQ5	776.72	1.17	0.37	0.000	0.320	0.554
local shimmer	36.47	15.31	0.47	0.000	0.008	0.508
shimmer APQ3	0.28	13.67	0.59	0.606	0.010	0.457
shimmer APQ11	107.1	7.35	0.03	0.000	0.035	0.865
HNR mn	96.56	23.84	1.39	0.000	0.003	0.261
intensity mn	26.06	2.1	0.07	0.000	0.198	0.799
H1-H2	112.07	7.46	3.42	0.000	0.034	0.089

3.3. Perturbation

With regard to local jitter, the Korean speakers show clear preferences for smaller values in the polite condition (14 of 16 had smaller values; 2 had significant differences with $p<0.05$). With respect to the RAP values, there was a more equivocal tendency for smaller values (6 higher out of 16 speakers; 2 significant with $p<0.05$) as well as for PPQ5 (7 of 16 speakers; 2 significant with $p<0.05$). There are also differences in local jitter (F1=11.31, p1<0.001; F2=3.12, p2=0.128) and near-significant differences in PPQ5 (F1=4.51, p1=0.052; F2=1.17, p2=0.32). We also observe strong gender effects (see Table 1). The German females show slightly higher local jitter values (6 out of 9), and PPQ5 (6 out of 9); although these differences are after all very subtle. Hence the ANOVA reveals no significant differences for all three jitter parameters. All measures of amplitude perturbation show the same preference: 14 of 16 Korean speakers had smaller local shimmer values, 13 of 16 had smaller APQ3 values and 14 of 16 had smaller APQ11 values in the polite condition. This is also supported by the results of the ANOVAs which show significant differences for local shimmer (F1=30.77, p1<0.001; F2=15.31, p2=0.008), APQ3 (F1=18.88, p1=0.0012; F2=13.67, p2=0.010) and APQ11 (F1=24.82, p1<0.001; F2=7.35, p2=0.035), but no gender effect for APQ3 (see Table 1). The German female speakers also show a preference for smaller shimmer values in APQ11 (8 of 9) and local shimmer (5 of 9), but the opposite tendency for APQ3 (2 of 9). For the German speakers, none of these differences comes out significant by means of the ANOVA.

Regarding noisiness, we observe highly significant differences between the two politeness registers for both HNR (F1=56.48, p1<0.001; F2=23.84, p2=0.003) and H1-H2 (F1=17.58, p1<0.001; F2=7.46, p2=0.034). For these parameters, we also obtained a highly significant gender effect (p1=0.04, p2<0.001): female participants increased breathiness in the polite condition to a much larger extent than male participants (see Table 1). This is in line with the finding that women have overall breathier voices [7]: It might be the case that "breathiness" is not employed by male speakers to indicate politeness because of its association with femininity.

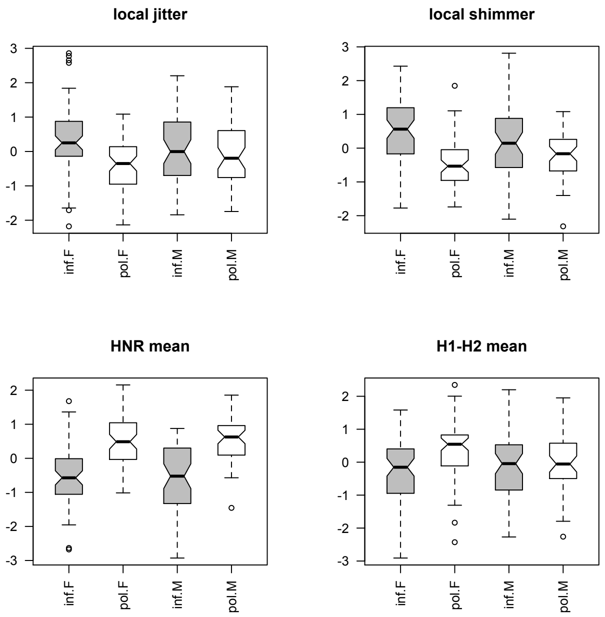

Figure 2: *local jitter, local shimmer, mean HNR and mean H1-H2 as z-scored values grouped by gender (F=female, M=male) and mode (inf=informal, pol=polite); Korean*

Table 2: *summary of results of participant-wise (F1(1,8)) and task-wise (F2(1,6)) ANOVA of the 9 female German speakers*

parameter	F1	p1	F2	p2
f0 median	16.38	0.0037	2.86	0.1417
period SD	4.36	0.0701	1.36	0.2882
local jitter	1.43	0.2659	0.18	0.6847
jitter RAP	2.62	0.1444	0.22	0.6546
jitter PPQ5	3.73	0.0896	0.15	0.7092
local shimmer	0.03	0.8754	0.06	0.8212
shimmer APQ3	3.24	0.1097	4.46	0.0791
shimmer APQ11	7.86	0.0231	2.37	0.1747
HNR mn	0.08	0.7899	0.01	0.9163
intensity mn	6.16	0.0380	2.13	0.1945
H1-H2	0.25	0.6299	4.91	0.0687

3.4. Speech Rate, Fillers & Pauses

For the Korean speakers, speech rate as measured by words per second differs significantly between the politeness registers: when speaking politely, the speech rate is slower than when speaking in the informal register (F1(1,14)=18.048, p1=0.001; F2(1,8)=56.709, p2<0.001). However, speech rate as measured by syllable rate (F1(1,14)=0.333, p1=0.573; F2(1,8)=1.540, p2=0.25) and articulation rate (F1(1,14)=0.013, p=0.91; F2(1,8)=0.699, p=0.427) shows no significant differences between the politeness registers. When speaking politely, there was a marked increase of laterally produced breath intakes (p1=0.029, p2=0.042). Also, there were more filled pauses such as *hm* or *ah* (p1=0.012, p2=0.025). The rate of unfilled pauses shows no significant effect. It thus appears to be the case that in Korean, audible hesitation markers are more likely to have social meaning than non-audible speech pauses. It also appears to be the case that the audible breath intake serves two functions (1) to fulfill the physiological requirements of the intake of air and (2) to express emotion and attitude (cf. Yuan & Li [14]). Parallel to e.g. laughter [13] we assume a patternized and learned control of these paralinguistic elements. We think that the higher amount of fillers and hesitation markers in the polite condition serves as a stylized way to mark insecurity.

3.5. NAQ

For the current NAQ analysis thus far, we have only considered data from 3 Korean female and 3 Korean male speakers. The overall tendencies suggest a higher NAQ for polite speech, and since higher NAQ values indicate relatively more breathiness, this is in line with our results of the HNR and H1-H2 analyses. However, one should note that 2 individuals seem to apply the opposite strategy and that so far, the ANOVAs did not real significant effects of politeness.

4. Discussion

We identified several acoustic and non-acoustic correlates of politeness. These results can be compared with studies focussing on the phonetic correlates of emotional expression (cf. e.g. [12]). In our data, the phonetic profile of politeness seems to be somewhat similar to the phonetic profiles of arousal/stress and fear/panic. This is also reflected in the increase of hesitation markers. We also observe that a number of measures of variability (jitter, shimmer, pitch variability) are decreased in the polite register, thus making Korean polite speech sound more “monotonous” and “streamlined”. It remains to be tested whether different speech act types have different prosodic profiles and also, to what extent the phonetic profile of politeness differs between speakers of different age groups. The fact that the German speakers behaved similar to the Koreans with respect to a number of phonetic parameters (f0, f0 variability, loudness, shimmer, jitter) points to possible cross-linguistic tendencies of how politeness is phonetically implemented. In this regard, it is also noteworthy that we conducted a forced-choice perception study with 7 English speakers who listened to pairs of polite and informal sentences taken from similar contexts. Even though the English speakers had no knowledge of Korean, they correctly identified the politeness level above chance level with about 60% accuracy ($t(6)=22.26$, $p<0.001$).

5. Conclusions

This study provides evidence for a phonetic difference between *contaymal* and *panmal*, a finding which extends to similar differences in German politeness levels. The chosen acoustic parameters serve to characterize long speech sequences which are spoken with the same “attitude” or in the same interactional context. Both Ohala [8] and Brown & Levinson [4, 267-268] predict a cross-linguistic association of sustained high pitch with politeness. However, these “universals” are better understood as cross-linguistic tendencies. Our data also show effects which stand against these prior claims. The prediction does not hold for Korean, German and American English [11].

In future work, we aim to control for the particular speech acts performed in the individual tasks and we want to see how phonetic differences between different speech acts relate to proposed theoretical notions of politeness (e.g. *negative* politeness-as-deference versus *positive* politeness-as-sympathy). Furthermore, we will extend this work to other languages (e.g. Russian or Arabic) and perform more cross-linguistic perception studies to find possible universals of how politeness is realized in phonetic and prosodic terms.

6. Acknowledgements

The authors would like to thank all Korean and German subjects for their participation, the MPI EVA (Bernard Comrie) and the Phonetics Department of Cologne University (Martine Grice, Anne Hermes) for hosting the recordings, Oh Ju-Young and Kim Won-Hee for their indispensable help with translations and instructions; and Benjamin Bergen (University of San Diego), Roger Mundry (MPI EVA) and Katie Drager (University of Hawai'i) for helpful comments and suggestions. Additionally we thank the three anonymous reviewers for their valuable comments.

7. References

[1] P.Alku and T. Bäckström, “Normalized amplitude quotient for parametrization of the glottal flow”. *The J. Acoust. Soc. Am.*, 112 (2): 701-710, 2002.

[2] P. Boersma, and D. Weenink, “Praat: doing phonetics by computer (version 5.1.17)”, [computer program], Retrieved 22 september, 2009, from http://www.praat.org. 2009.

[3] D. Bolinger, “Intonation across languages”. In J. Greenberg, editor, *Universals of human language*, 471–524. Stanford UP, Stanford, 1978.

[4] P. Brown and S. Levinson, *Politeness : Some Universals in Language Usage*. Cambridge University Press, 1987.

[5] N. Campbell and P. Mokhtari, “Voice quality: the 4th prosodic dimension”, *15th ICPhS, Barcelona*, 2417–2420, 2003.

[6] M. Ito, “Politeness and voice quality-the alternative method to measure aspiration noise”. In *Speech Prosody 2004, International Conference of, Nara, Japan*. ISCA, 2004.

[7] D. Klatt and L. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers”. *Journal of the Acoustical Society of America*, 87(2):820–857, 1990.

[8] J. Ohala, “An ethological perspective on common cross-language utilization of F0 of voice”, *Phonetica*, 41: 1-16, 1984.

[9] Y. Ohara, “Finding one's voice in japanese: A study of the pitch levels of l2 users”. In A. Pavlenko and A. Blackledge, editors, *Multilingualism, second language learning, and gender*, 231–254. Walter de Gruyter, 2001.

[10] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2.10.0 edition, 2009.

[11] S. Shin, *Grammaticalization of politeness: A contrastive study of German, English and Korean*. PhD thesis, University Of California, Berkeley, 2005.

[12] K. Scherer, “Vocal communication of emotion: A review of research paradigms”, *Speech communication*, 40(1-2):227–256, 2003.

[13] J. Trouvain, “Segmenting phonetic units in laughter”. In *Proc. 15th International Conference of the Phonetic Sciences, Barcelona, Spain*, 2793–2796, 2003.

[14] C. Yuan and A. Li, “The breath segment in expressive speech”. *Computational Linguistics and Chinese Language Processing*, 12(1):17–31, March 2007.