Journal of Phonetics 40 (2012) 808–815

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

The phonetic profile of Korean formal and informal speech registers

Bodo Winter [a,*], Sven Grawunder [b]

[a] Department of Cognitive and Information Sciences, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA
[b] Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

ARTICLE INFO

Article history:
Received 2 October 2011
Received in revised form
23 July 2012
Accepted 27 August 2012
Available online 3 October 2012

ABSTRACT

In this exploratory sociophonetic study, we investigated the properties of formal and informal speech registers in Korean. We found that in formal speech, Korean male and female speakers lowered their average fundamental frequency and pitch range. The acoustic signal furthermore exhibited overall less variability, as evidenced by decreased fundamental frequency and intensity standard deviations, and decreased period and amplitude perturbations. Differences in speech registers affected Harmonics-to-Noise-ratio and the difference between the first and second harmonic as well, suggesting breathiness-related changes, and the speech was slower and included more non-lexical fillers such as *ah* and *oh*. Unexpectedly, formality also affected breathing patterns, leading to a noticeable increase in the amount of loud ‘‘hissing’’ breath intakes in formal speech. We thus show that a variety of different means of vocal expression play a role in signaling formality in Korean. Further, we outline the implications of this study for phonetic theory and discuss our results with respect to the Frequency Code and research on clear speech.

Published by Elsevier Ltd.

1. Introduction

In everyday interactions, it is often crucial that we choose the appropriate level of formality in order to communicate effectively. This means not only choosing the ‘‘right words’’ but also choosing the right way of saying them (Ng & Reid, 2001, p. 362). The present study addresses the question of how this ‘‘right way of saying things’’ can be characterized, and how phonetic patterns of formality impact phonetic theory.

Whereas research on the expressive dimension of speech is gaining momentum (Erickson, 2005; Scherer, 2003; Tatham & Morton, 2004), sociopragmatic aspects such as the phonetics of formal language have received relatively little attention. Hence, not much is known about the vocal strategies speakers use when speaking formally. This study sets out to fill this gap. We conducted a speech production task with 16 native speakers of Korean who spoke utterances in either a formal or an informal speech register.

Korean was chosen because this language is widely known for its elaborate system of honorification (see e.g. Sohn, 1999) that has received a lot of attention from linguists (e.g. Shin, 2005; Yoon, 2004). The well-defined morphosyntactic and lexical differences between the Korean speech registers facilitate the task of eliciting formal versus informal language in a laboratory setting. The Korean formal speech register *has to be used* when speaking to strangers or superiors and it is thus a dominant feature of the language that is employed on a daily basis. Formal speech in Korean has been characterized as a normative form of politeness (e.g. Byon, 2006; Sohn, 1999), rather than ‘‘volitional politeness’’ (Sohn, 1999, pp. 407–418). We thus relate formality to the already existing body of work on phonetic politeness.

The study of phonetic correlates of formality is important for several reasons. First, it is important for sociophonetics. Most sociophonetic investigations have focused on individual units such as specific intonation contours (e.g. Podesva, 2011) or the socioeconomic and dialectal variation of allophones (e.g. Docherty & Foulkes, 2005). This study, on the other hand, examines the broad characteristics of whole utterances, or the global as opposed to the local properties of speech. We show that these vary systematically along known sociophonetic dimensions such as the status of the interlocutor. These global changes are best characterized with Tatham and Morton's (2004) speech model where expressive and sociopragmatic factors are part of an ‘‘expressive envelope’’ that affects all phonetic units within an utterance, ranging from individual segments to intonation contours.

Another important reason to study formal language is because of intricate connections to biological codes of speech (Gussenhoven, 2004, chap. 5). Given that formal speech is used with superiors, and is often analyzed as a type of politeness (for Korean see Byon, 2006; Sohn, 1999), this has invited the idea that it might be connected to Ohala's Frequency Code (Ohala, 1984, 1994). This hypothesis seeks to unify a number of seemingly unrelated phonetic and phonological facts (e.g. question

* Corresponding author. Tel.: +1 559 799 5517.
E-mail address: bodo@bodowinter.com (B. Winter).

0095-4470/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.wocn.2012.08.006

intonation, sound symbolism) in a common biological framework. A central component of this hypothesis is that low-pitched speech signals dominance, and high-pitched speech signals subdominance. Many researchers have found or suggested that there is an association between high pitch and polite speech (e.g. Brown & Levinson, 1987, pp. 267–268; Ohala, 1984, 1994; Pike, 1945, p. 59), suggesting that the vocal expression of subdominance could be mapped onto politeness. For example, Tzeltal speakers employ a "sustained falsetto" in polite greeting formulae, while in Tamil, low-caste speakers address high-caste speakers in a "customary high-pitched voice" (Brown & Levinson, 1987, p. 267). Irvine (1979, p. 777) argues that in Wolof, high pitch suggests low social rank, and low pitch high social rank. Loveday (1981) and Ohara (2001) found that female speakers of Japanese tended to express politeness by raising their average f0, and Ofuka, McKeown, Waterman, and Roach (2000) showed that utterance-final pitch rises in Japanese were rated as being more polite than final falls. However, it has been claimed that in Korean, female speakers *lower* their pitch when speaking to superiors (Shin, 2005), thus rendering the language an exception to this proposed cross-linguistic pattern. By investigating Korean in more detail, we seek to shed light on the nature of this exception.

The study of formal language can also inform considerations of another biological code, the Effort Code. According to this hypothesis, increasing speech production effort leads to greater articulatory precision and pitch expansion (Gussenhoven, 2004). Clear speech often leads to an expansion of the pitch range and an increase of f0 (Smiljanic & Bradlow, 2005, 2008), but no such pattern has been observed in Korean (Cho, Lee, & Kim, 2011). The present study investigates to what extent Korean formal speech mirrors previous findings on clear speech, thus pointing to connections between clarity and formality.

In this study, we investigate a whole range of phonetic correlates of formality. Our choice of phonetic measurements is driven by prior research on politeness. Besides the above-mentioned studies investigating f0, voice quality has been suggested to be of importance. Ito (2004) found an increase in the perturbation of aspiration noise to be associated with politeness in Japanese. And, based on measurements of the normalized amplitude quotient of a single female Japanese speaker, Campbell (2004) found a more breathy phonation to be associated with talking to strangers as opposed to family members and friends. Ofuka (1996) and Ofuka et al. (2000) investigated speaking rate and found that medium levels of speaking rate were judged to be more polite than extremely slow or fast rates. To our knowledge there is currently no work that has looked at pauses, fillers or breath intakes with respect to politeness or formality. We nevertheless include these measurements because they have been shown to be of general pragmatic importance (e.g. Clark & Fox Tree, 2002).

In investigating the phonetic correlates of formality, we thus pursued an integrative and holistic approach that combines a number of different phonetic measurements. The few previous studies that are relevant to this topic have mostly focused on a small set of measures. Taken together, the literature covers a range of different phonetic correlates, suggesting that sociopragmatic aspects affect the speech signal at a very general level. However, previous studies were usually based on analyses of different data sets; in this work, we intend to show the influence of formal vs. informal speech registers on various phonetic measures in the same dataset.

2. Methodology

2.1. Participants

Nine female and seven male native speakers of Korean participated in our study (age: 21–31, median age: 23.5). Of the sixteen speakers, all but three were from the Seoul metropolitan area. At the time of the recording, the participants had resided in Germany for a mean time of four years and all but one reported to use Korean on a daily basis. None of the participants reported to have any problems with hearing, reading or eyesight.

2.2. Procedure

Participants were instructed by a native speaker of Korean, as well as via Korean instructions on a computer screen. Participants were told that the study was about formality. After reading out a short Korean newspaper article to adjust the sound level and to get speakers acquainted with the recording situation, each participant performed two different tasks. First, participants were given a note in paper format. They had to use the main points of this note (e.g. "meeting at 10.30 am in front of Starbucks") to formulate a coherent message which they had to leave on an imaginary cell phone mailbox (this "Mailbox Task" has also been used by Shin, 2005). Second, a verbal version of the Discourse Completion Task (cf. Byon, 2006) was performed in which participants were given written context passages, e.g. "You are in the professor's office and want to ask for a letter of recommendation". The participants then responded by formulating the first utterance of a role-played dialog. As soon as the participants read and understood a context passage in each of the tasks, a picture of the interlocutor appeared. After two seconds delay, participants heard a beep which served as a signal to deliver a response.

2.3. Stimulus materials

In total, there were two stimuli pairs (formal and informal) for the Mailbox Task and five stimuli pairs for the Discourse Completion Task. We used a range of different interaction types such as "making an appointment", "asking for a favor" (e.g. a letter of recommendation) and "apologizing for coming too late". Although differences between interaction types are expected to occur, for the current study we only aim at characterizing formal language that co-occurs with the use of honorifics across different interaction types. Formality was operationally defined as the difference between the cultural categories *contaymal* (formal speech) and *panmal* (informal speech).[1] Speakers of Korean generally use *contaymal* when speaking to superiors and strangers, whereas they use *panmal* when speaking to peers and inferiors. The two registers differ morphosyntactically along a number of dimensions, ranging from pronouns to verbal and nominal suffixes (Sohn, 1999).

We chose context passages with male interlocutors that Korean speakers would clearly address in either formal speech or informal speech, for example a professor or a CEO versus a same-aged room-mate or a friend. The interlocutors differed along a number of different dimensions: age, occupation, social distance and social power.[2] The pictures of the interlocutors that were presented to our participants emphasized the power and age differences.

[1] The fact that we focus our analysis on a two-fold distinction between these registers does not imply that the registers are stable and unchanging. We leave it for future studies to investigate what effect different speech acts and interaction types within these basic categories have.

[2] Manipulating power was crucial because power differences between speaker and interlocutor reliably predict the degree of politeness in a number of diverse communicative contexts (cf. Holtgraves, 2001, p. 348), and power and distance may be among the two most important dimensions of social interaction (Wish, Deutsch, & Kaplan, 1976). Yoon (2004) argues that for Korean honorifics, the most important distinction is between interlocutors who are "above" the speaker versus those who are "below".

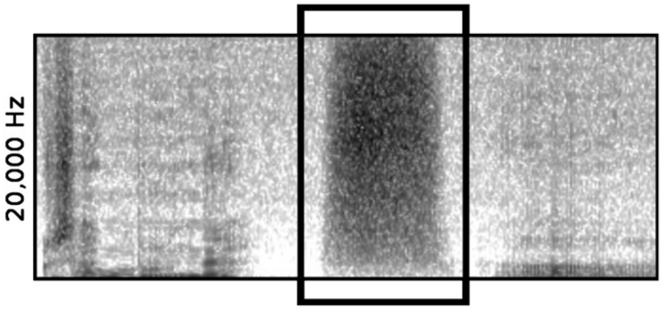

Fig. 1. Sonogram showing a representative loud noisy breath intake (marked by box) that exhibits high-frequency noise components up to 20,000 Hz. The breath intake is 451 ms long, the whole sequence is 2.15 s. Note that the breath intake is surrounded by short pauses and followed by a filler [aː].

2.4. Recordings

All speakers were recorded in a sound-attenuated booth at the Institute of Phonetics in Cologne with a head-set microphone AKG C420 (linear characteristic) with 48 kHz/16 bit sampling. The distance (approx. 5 cm) and orientation of the participants to the microphone, as well as the input level of the sound recording, was held constant across all recording sessions. In total, we collected and analyzed a total of 1 h and 30 min across all participants (this excludes the breaks between the different stimuli).

2.5. Phonetic analyses

Phonetic analyses were conducted with Praat 5.2.16 (Boersma & Weenink, 2011). We first manually labeled all silent pauses, audible breath intakes, fillers such as *ah* and *oh* (henceforth "oral fillers") and nasal fillers such as *mh* or *nh*. Our speakers furthermore frequently produced noticeably loud breath intakes (henceforth "noisy breath intake"). These had the auditory quality of a retracted sometimes bidental central fricative, but they often involved a more lateral place of articulation, akin to a lateral fricative. These ingressive sounds have a "hissing" perceptual quality and a strong high-frequency component, sometimes going up to 20 kHz. Like interjections, they were often detached from the syntactic environment and surrounded by pauses, which made them very prominent (see Fig. 1). This noisy breath intake seems to be reminiscent of the ingressive "hiss" in Japanese (Critchley, 1939) or Chinese (Eklund, 2008, after Key, 1975, p. 69).

For our pause measurements, silence exceeding 200 ms was counted as pause. This is different from the common 100 ms threshold (e.g. Butcher, 1981; Trouvain, 1999; Trouvain & Grice, 1999) because in Korean, many stop closure durations exceed 100 ms. The labeling procedure resulted in a total of 2497 pauses and non-lexical elements. Of these, 150 items (6%) were excluded because they occurred in the context of disfluencies (restarts, slips of the tongue, within-word pauses exceeding 200 ms and silent pauses exceeding 2.5 s were counted as disfluency). For articulation rate measurements, we counted syllables as they appeared in the transcripts transcribed by a native speaker, and divided these counts by the duration of each response. As opposed to speaking rate, articulation rate refers to the number of syllables per second *excluding* pauses and fillers (cf. Trouvain, 2004).

For f0 and voice quality measurements, all voiced portions were extracted from the non-pause and non-filler speech parts and concatenated into a single stream. We calculated pitch using the forward cross-correlation algorithm (Boersma & Weenink, 2011; Talkin, 1995) with slightly altered default settings in Praat ('To Pitch… 0 75 15 yes 0.03 0.55 0.01 0.35 0.14 600'). F0 range was calculated by taking the difference between the 95th and the 5th percentile of all of the f0 values of an utterance. Intensity was measured in dB SPL; the intensity range was similarily taken to be the difference of percentiles.

We calculated local jitter (absolute period-to-period difference divided by the average period) and local shimmer (likewise normalized amplitude difference of consecutive periods). Based on the pitch estimation above we used Praat default settings, with a 0.0001 s period floor and 0.02 s period ceiling. For jitter measurements, the maximum difference factor was 1.3 between consecutive intervals, and for shimmer measurements, the difference factor was 1.6 for consecutive amplitudes. Harmonics-to-Noise ratio (HNR) was calculated by means of the cross-correlation pitch algorithm with a 0.01 s time step, 0.013 window length, 0.1 frame silence threshold and 1 period per window. We calculated the difference between first and second harmonics ($H1-H2$), a possible index of breathiness (e.g. Hillenbrand & Houde, 1996; Shrivastav & Sapienza, 2003). We report $H1^*-H2^*$, a corrected measure that takes differences in formant positions into account by using the approach proposed by Iseli and Alwan (2004), provided by the Voicesauce software package (Shue, Keating, Vicenik, & Yu, 2011). This measurement was based on the central vowel portion of each vowel.

2.6. Statistical analyses

All data were analyzed using R (R Development Core Team, 2009) with the R packages *lme4* (Bates, Maechler, & Bolker, 2011), *languageR* (Baayen, 2009; cf. Baayen, 2008) and *glmmADMB* (Skaug, Fournier, Nielsen, Magnusson, & Bolker, 2012). We performed a series of generalized linear mixed effects models (GLMMs) with the within-subject fixed effect Speech Register (formal vs. informal) as a test variable, and the between-subject fixed effect Gender (male vs. female) as a control variable. If there was a significant Speech Register $\times$ Gender interaction, *p*-values will only be reported for the interaction effect (because the significance of main effects is uninterpretable in case of a significant interaction, see e.g. Zar, 1999). For statistical details, see Appendix A.

3. Results

In general, participants adapted very well to the task and used the appropriate morphosyntactic forms for the different registers. In line with this, the informal interjection *ya* [jaː] appeared almost exclusively in the informal condition (98%), and the polite *yey* [jeː] almost exclusively in the formal one (96%). These general measures show that the elicited speech exhibited common features of the Korean formal and informal registers.

Fig. 2 shows boxplots of the mean f0 values and the f0 range for formal and informal speech. Comparing the two conditions, the mean f0 was on average 17 Hz $\pm$ 5.4 Hz lower in formal speech ($p=0.005$). Fundamental frequency variability was 0.42 $\pm$ 0.13 Hz standard deviations lower in formal speech ($p=0.001$), and the f0 range was lowered by 1.43 $\pm$ 0.56 semitones ($p=0.006$), see Fig. 2b. None of the f0 measurements produced significant interaction effects with the factor Gender, showing that the attitude-specific use of f0 was consistent across males and females.

For average intensity, there was neither a main effect of Speech Register ($p=0.5$) nor a Speech Register $\times$ Gender interaction ($p=0.14$). There were interaction effects for variability in intensity ($p=0.007$) and intensity range ($p=0.005$). The intensity variability was lower in the formal condition by 0.72 $\pm$ 0.15 dB standard deviations, however only for females (interaction

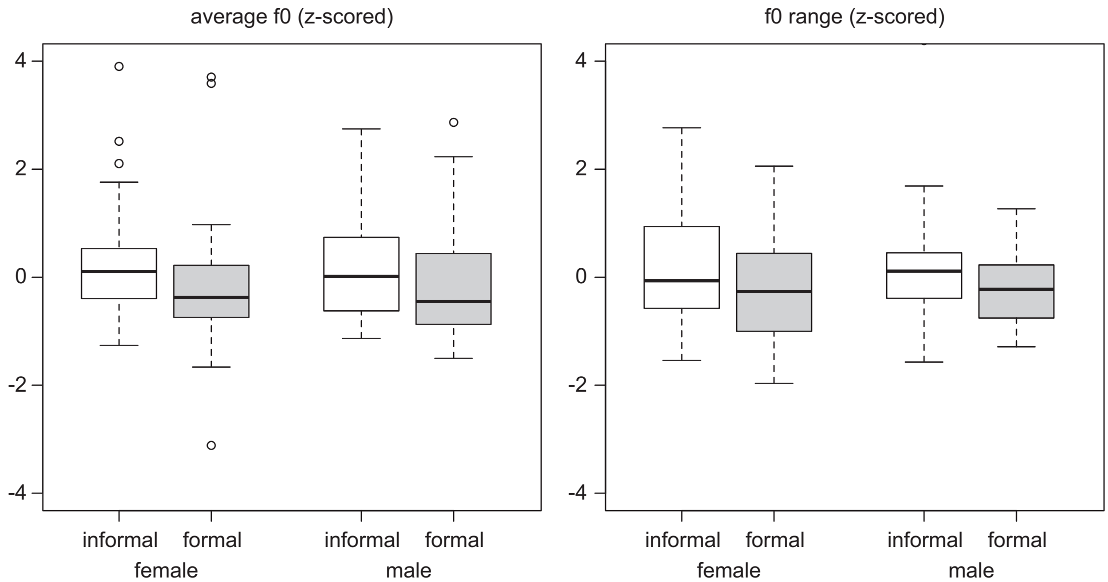

Fig. 2. Utterance-level mean f0 and f0 range, *z*-scored to a mean of 0 and a standard deviation of 1. Each box covers 50% of the respective data. The median is indicated by bold lines. Whiskers indicate 1.5 times the interquartile range.

coefficient; for males in the formal condition: $+0.65 \pm 0.23$ dB). The intensity range was also lower, by about 2.55 ± 0.54 dB. This effect was again driven by females (interaction: $+2.43 \pm 0.82$); males produced almost no difference between formal and informal speech with respect to intensity range.

Local jitter (overall average: 0.019) was decreased in the formal condition by 0.0011 ± 0.0004 ($p = 0.015$), with no interaction. The effect of Speech Register came in the form of an interaction for local shimmer measurements ($p = 0.007$). Local shimmer (overall average: 0.066) was decreased in the formal condition (-0.0079 ± 0.0014), but much more so for females (interaction for male/formal: $+0.0062 \pm 0.0021$).

For HNR, there was a nearly significant Speech Register × Gender interaction ($p = 0.053$). HNR was increased in the formal condition by about 1.51 ± 0.19 dB, but it was less increased for males (interaction: -0.59 ± 0.29). For H1*−H2*, there was a significant interaction ($p = 0.03$), with a decrease by 0.66 ± 0.28 dB for the formal condition.[3] The positive coefficient for the interaction (1.0 ± 0.43 dB) indicates that males had the exact opposite pattern, with a slight increase of H1*−H2* values for the formal condition (by about 0.34 dB).

As measured by articulation rate, participants spoke on average 0.38 ± 0.18 syllables per second more slowly in the formal condition (main effect: $p = 0.04$). For the count of silent pauses, there was a significant Speech Register × Gender interaction ($p = 0.001$). There were overall less silent pauses in the formal condition[4] (-0.58 ± 0.14 average count), but this effect was driven by females (interaction for male/formal: $+0.47 \pm 0.18$). Pauses filled with oral fillers, such as *ah* and *oh*, were increased in the formal condition (0.36 ± 0.097, $p = 0.0002$), with no interaction. Pauses filled with nasal fillers such as *mh* and *nh* exhibited an interaction ($p = 0.008$): while there was an overall increase of 0.52 ± 0.58 fillers, males showed the opposite pattern

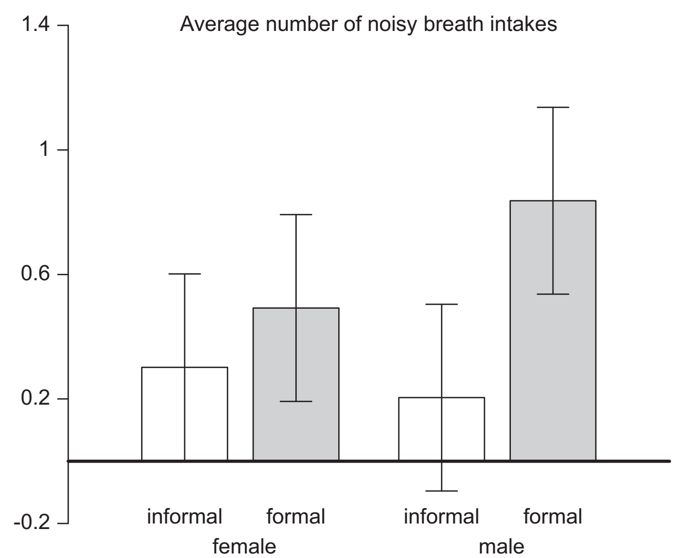

Fig. 3. Average number of noisy breath intakes. Bars indicate standard errors taken from the statistical model discussed in the paper.

(interaction: -1.74 ± 0.68). Females used nasal fillers overall less than males and they did not show much of a difference between the formal and informal condition. Males on the other hand reduced their use of nasal fillers when speaking formally. These results show that pauses filled with nasal fillers and pauses filled with oral fillers behave differently with respect to formal speech.

While regular breath intakes showed no main effect ($p = 0.2$) or interaction effect ($p = 0.74$), noisy breath intakes ("hisses") produced a Speech Register × Gender interaction ($p = 0.03$), depicted in Fig. 3. Overall, there were more noisy breath intakes in the formal condition (0.3 ± 0.3), but even more so for males (1.0 ± 0.46). While females had, on average, 1.63 times more "hisses" in the formal than in the informal condition, males had about 4.1 times more when speaking formally.

Table 1 sums up all results and highlights that for the 16 different measures that we investigated with respect to formality,

[3] These results differ from Grawunder and Winter (2010), where the more unreliable uncorrected H1−H2 values were used and females had higher H1−H2 values in the formal condition.

[4] We did not obtain an effect in Grawunder and Winter (2010), presumably because our previous pause counts did not differentiate between completely silent pauses and pauses filled with breath intakes.

Table 1
List of investigated phonetic parameters with coefficients, standard errors and *p*-values from the statistical models described in the paper. Main effect coefficients indicate the change from the informal to the formal category. Interaction coefficients are the changes for males in the formal condition. "n.s." indicates "non significant". In case of a significant interaction, the significance of main effects is uninterpretable and set to N/A. Square brackets indicate total counts of the respective item. The column "Any effect?" shows that for all but two measures, there was an effect of Speech Register (as a main effect or an interaction effect).

Phonetic parameter	Estimate	Main effect	Estimate	Interaction	Any effect?
f0 (Hz)	-17.2 ± 5.4	$p=0.005$	5.56 ± 8.24	n.s.	Yes
f0 SD (Hz)	-0.42 ± 0.13	$p=0.001$	0.27 ± 0.2	n.s.	Yes
f0 range (semitones)	-1.43 ± 0.56	$p=0.006$	-0.13 ± 0.84	n.s.	Yes
Intensity (dB)	-0.25 ± 0.36	n.s.	-0.84 ± 0.54	n.s.	**No**
Intensity SD (dB)	-0.72 ± 0.15	N/A	0.65 ± 0.23	$p=0.007$	Yes
Intensity range (dB)	-2.55 ± 0.54	N/A	2.43 ± 0.82	$p=0.005$	Yes
Local jitter	-0.0011 ± 0.0004	$p=0.015$	-0.00004 ± 0.0007	n.s.	Yes
Local shimmer	-0.0079 ± 0.0014	N/A	0.0062 ± 0.0021	$p=0.007$	Yes
Harmonics-to-noise Ratio (dB)	1.51 ± 0.19	N/A	-0.59 ± 0.29	$p=0.053$	Yes
$H1^*-H2^*$ (dB)	-0.66 ± 0.28	N/A	1.0 ± 0.43	$p=0.03$	Yes
Articulation rate (syllables/second)	-0.38 ± 0.18	$p=0.04$	-0.06 ± 0.27	n.s.	Yes
Silent pause count [670]	-0.58 ± 0.14	N/A	0.47 ± 0.18	$p=0.001$	Yes
Oral filler count [474]	0.36 ± 0.097	$p=0.0002$	-0.25 ± 0.21	n.s	Yes
Nasal filler count [70]	0.52 ± 0.58	N/A	-1.74 ± 0.68	$p=0.008$	Yes
Regular breath intake count [1032]	0.16 ± 0.12	n.s.	-0.06 ± 0.17	n.s.	**No**
Noisy breath intake count [101]	0.3 ± 0.3	N/A	1.0 ± 0.46	$p=0.03$	Yes

all but two were significant. This means that for 14 of 16 dependent variables, there was some effect of formality, either as main effect or as interaction effect. Obtaining so many significant results when conducting 16 tests is highly unlikely overall (less than $p<0.0001$, Cross & Chaffin, 1982). There were 8 interaction effects and 6 main effects of Speech Register, showing that formality affected phonetic variables sometimes in a manner that was consistent across males and females, but slightly more often in a manner that was noticeably different for males and females.

4. Discussion

Contra to cross-linguistic predictions and in support of Shin (2005), we found that Korean speakers lowered their f0 when speaking in the formal speech register. This pattern is different from Japanese (Loveday, 1981; Ohara, 2001) and other languages, and it stands against the notion that speech directed to superiors should always be associated with high f0, as predicted by the Frequency Code (Ohala, 1984, 1994) and by Brown and Levinson (1987, pp. 267–268). However, the question arises as to whether biological codes such as the Frequency Code should apply to Korean formal language at all. As mentioned above, Korean honorifics (and associated phonetic patterns) are a normative form of social indexing that *have to be used* in certain contexts. Most researchers interested in politeness–pitch associations have not spent much time on defining what they mean by politeness, and in our case, where we are dealing more with formality than "volitional politeness", purported pitch–politeness associations might simply not apply. It should also be pointed out that previous investigations of pitch in relation to social status have only focused on a handful of languages (in particular Japanese) and comprised mostly anecdotal observations. Thus, whether there is indeed a cross-linguistic association between pitch and politeness, or between pitch and formality, is still an open question. Moreover, the generality of the Frequency Code needs to be interpreted carefully with respect to each phenomenon that it seeks to explain, as some of these phenomena might not necessarily be connected to the hypothesis. In our case, the use of a normative code might simply not be within the explanatory domain of the Frequency Code.

A possible reason for the lowering of fundamental frequency in the formal condition is that higher f0 levels might signal "animatedness" or arousal, whereas lower f0 levels are often used to signal a more neutral or sad mood (see Juslin & Laukka, 2003; Pell, 2001). It is interesting to note that the pattern of a "monotone" or "damped" vocal expression that we found – lower intensity, lower intensity variability, lower f0, lower f0 variability, slower articulation rate – lines up very well with what has been found on vocal correlates of a sad or neutral tone of voice (Johnstone & Scherer, 2000; Juslin & Laukka, 2003; Scherer, 2003). Similar to this, Irvine (1979, p. 780) mentions that in Mursi (an ethnic group in Ethiopia), formality is associated with an absence of "excitement", and more generally, a reduction in variation different cultures (Irvine, 1979, p. 774)

Interestingly, Cho et al. (2011) report that Korean does not increase average f0 in clear speech either, whereas other languages such as English and Croatian do (Smiljanic & Bradlow, 2005, 2008). Moreover, clear speech in Korean does not involve pitch range expansion (Cho et al., 2011), a feature of clear speech in other languages. Although we found a lowering of average f0 and f0 range in our data, whereas Cho et al. (2011) found no effects for these measurements in clear vs. casual speech, our data bear a certain resemblance to clear speech: The articulation rate was slowed down, and as indicated by $H1^*-H2^*$ and HNR values, the formal speech of our participants was less breathy and had stronger harmonic components, suggesting an overall less noisy, clearer signal. Intuitively, it seems plausible that speakers might be more careful with their choice of words and the way they construct and pronounce sentences in order not to insult superiors (see also Labov, 1972, p. 113, who links formality to increased attention), and this increased care is likely to have phonetic correlates. While we showed this for the broad characteristics of Korean, this should extend to the segmental domain as well. We would thus predict that segments should be more hyperarticulated in formal speech. At the least, our data shows that when phonetic studies investigate the effects of clear speech on such parameters as f0 and voice quality, the formality of an utterance has to be taken into account because it also affects these aspects of the speech signal.

Moving on to the non-lexical speech items, these also bear the imprint of increased care in sentence production. Fillers such as *ah* and *oh* were increased in the formal condition, and these might signal increased care taken in sentence planning (Clark & Fox Tree, 2002). Silent pauses, on the other hand, were *decreased* in the formal condition, possibly indicating that in formal speech towards superiors, Koreans are more likely to use fillers (which contain more ''phonetic material'' than silent pauses) to make the care in sentence planning more obvious.

Clark and Fox Tree call non-lexical items such as fillers and pauses "signals, not symptoms" (2002, p. 105). In our case, this signaling view of non-lexical elements even appears to extend to such seemingly non-social units of speech as breath intakes. Although there was no difference in regular breath intakes, there was a marked difference in loud noisy breath intakes. In our study, these apparently perform the dual function of fulfilling the physiological requirement of the intake of air and the expression of social meaning (cf. Yuan & Li, 2007). The difference in noisy breath intakes between the formal and the informal condition shows that an element of speech that has a very crucial function – breathing – can be modulated by the social context of an utterance. Moreover, as many of these noisy breath intakes had a lateral place of articulation, this shows that a sound that is not even part of regular Korean phonology can have sociophonetic significance. Interestingly, breath intakes are often excluded from analyses in pragmatic studies (e.g. Ward, 2006, p. 134), but our data show that even breathing patterns can correlate with socio-pragmatic categories in a systematic fashion. The fact that *loud* breath intakes were more frequent in formal speech, and that the more open oral fillers were more frequent (compared to the decrease in nasal fillers) might be interpreted more generally in terms of the Effort code, in that the articulations that incur more effort in terms of sonority or subglottal pressure seem to be characteristic of the formal speech register.

Finally, one comment needs to be made about the size of the obtained effects. While some differences between the formal and informal registers were large (e.g. especially in terms of the rate of noisy breath intakes, fillers, nasal elements), most of the differences were very small (all intensity measurements, f0 range, articulation rate). Thus, the fact that there are multiple cues for formal speech seems to be important for signaling registers in a robust way. Some individual cues might be weaker, but by virtue of having many different cues, formality-related differences in speech can still be signaled efficiently. This mirrors the situation with phonological contrasts, where there is usually an abundance of cues (Hawkins, 2010; Winter & Christiansen, 2012), and it mirrors the cue layering present in other domains of speech such as cues for word boundaries (Christiansen, Allen, & Seidenberg, 1998). In our case, we also found a multiplicity of cues, signaling, however, a paralinguistic or pragmatic contrast rather than a purely linguistic one. Given the large quantity of cues, the perception of formal versus informal speech registers should be possible on the basis of the speech signal alone.

5. Conclusions

Previous work on the phonetics of polite and formal speech has only focused on a small number of phonetic features, primarily f0. Our data show that there are not just pitch differences between the formal and informal registers in Korean but also differences in intensity, $H1^*-H2^*$, HNR, jitter and shimmer, alongside differences in articulation rate and the rate of pauses, fillers and non-lexical elements such as breath intakes. We found significant effects for nearly every parameter that we investigated, which highlights that prior research has perhaps given a disproportionate amount of attention to the phonetic feature f0, while neglecting voice quality and non-lexical speech elements. Future work needs to test the perceptual relevance of the correlates we found and the degree to which the different phonetic correlates interact with each other in perception. Moreover, differences between interaction types and speech acts within the formal and informal speech registers will provide a fruitful topic of future research.

In terms of the biological codes discussed in the phonetic literature, we noted that our data do not have to be interpreted in terms of the Frequency Code and that overall, it is questionable whether this code necessarily applies to formality. Some aspects of our data are more readily interpreted in terms of the Effort Code, as formal speech in Korean shares some similarities with clear speech in Korean, and as it exhibits signs of increased care in sentence planning, as well as potentially more ''effortful'' fillers. Furthermore, to determine how formal speech and clear speech are related, one needs to compare these registers within speakers. Based on our results we can form the expectation that phonetic patterns of clear speech should be somewhat correlated with phonetic patterns of formal speech from the same speakers, and they should both be very different from informal and non-clear speech.

Linguistic research has emphasized morphosyntactic or lexical markers of formality. Our study shows that alongside these verbal markers, there also is a range of phonetic ones. Thus, the speech signal alone – independent of the choice of words or grammatical forms – conveys some of the social meaning of an utterance (cf. Tatham & Morton, 2004). By looking at the phonetic characteristics of formal speech, we can complement previous work on the morphosyntactic and lexical aspects of formal language, and we gain a more complete picture of the different dimensions that play a role for sociopragmatics. Finally, we can begin to understand the important relations between formality and such phonetic topics such as clear speech and the biological codes of speech production.

Acknowledgments

This work was in part supported by Uhde, ThyssenKrupp Technologies. This work has benefited a lot from the feedback we received for presentations at the *DGfS 2009*, the *AKSE 2009*, the *Japanese/Korean Linguistics 2009*, *Speech Prosody 2010* and the Max Planck Institute for Evolutionary Anthropology. We thank Daniel Voigt, Katie Drager, Bernard Comrie, Amy Schafer, Clive Winter and the editors and reviewers of *Journal of Phonetics* for useful comments and suggestions. We also thank Roger Mundry for statistical advice. We thank Oh Ju-Young and Kim Won-Hee for translations, transcriptions and helping out during the recording sessions and participant acquisition. We thank the Institute of Phonetics at the University of Cologne for making their recording facilities available to us. Most of all, we thank our participants, without whom this project would not have been possible. Only the authors are responsible for all remaining errors.

Appendix A

To avoid the language-as-fixed-effect fallacy (Clark, 1973) and account for non-independence (Winter, 2011), we used both subjects and items as crossed random effects (see Baayen, Davidson, & Bates, 2008). Likelihood ratio tests showed that random slope models (subject-specific slopes for the fixed effect Speech Register) were not necessary for any dependent measure, so we constructed random intercept models. For all count data, we initially constructed poisson models, but due to overdispersion (larger variance than the mean), we decided to use negative binomial models (Ismail & Jemain, 2007). With these models, we accounted for effort (duration of a response) and excess zeros (by setting zeroInflation=TRUE with glmmADMB). For all other variables we used Gaussian error distributions. For the Gaussian models, plots of residuals against fitted values revealed no obvious deviations from normality or homogeneity. For all models, we also evaluated DFBeta, which indicated that there were no overly influential cases.

All models were first evaluated with likelihood ratio tests (test model vs. null model with only the control variable Gender). If the full model vs. null model comparison reached significance, we calculated Markov Chain Monte Carlo-estimated p-values for the different fixed effects (Baayen et al., 2008). For the count variables, we present p-values based on likelihood ratio tests. Given the lack of degrees of freedom with mixed models, we refrain from reporting *df*. All results will be reported with standard errors. To control for multiple testing, we used Cross and Chaffin's (1982) binomial approach: the overall probability of obtaining as many significant results as we did turned out to be below $p < 0.0001$.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2009). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 0.955.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42.

Boersma, P., & Weenink, D. (2011). *Praat: Doing phonetics by computer* (Version 5.1.05) [Computer program]. Retrieved April 4, 2011, from 〈http://www.praat.org/〉.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. In Kohler, K. (Ed.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung*, 15.

Byon, A. S. (2006). The role of linguistic indirectness and honorifics in achieving linguistic politeness in Korean requests. *Journal of Politeness Research*, *2*, 247–276.

Campbell, N. (2004). Accounting for voice-quality variation. In *Proceedings of the second international conference on speech prosody* (pp. 217–220). Nara.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*, 344–361.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.

Critchley, M. (1939). Spastic dysphonia ("inspiratory speech"). *Brain*, *62*, 96–103.

Cross, E. M., & Chaffin, W. W. (1982). Use of the binomial theorem in interpreting results of multiple tests of significance. *Educational and Psychological Measurement*, *42*, 25–34.

Docherty, G. J., & Foulkes, P. (2005). Glottal variants of /t/ in the Tyneside variety of English: An acoustic profiling study. In: W. Hardcastle, & J. Mackenzie Beck (Eds.), *A figure of speech—A Festschrift for John Laver* (pp. 173–199). London: Lawrence Erlbaum.

Eklund, R. (2008). Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association*, *38*, 235–324.

Erickson, D. (2005). Expressive speech: Production, perception and application to speech synthesis. *Acoustical Science and Technology*, *26*, 317–325.

Grawunder, S., & Winter, B. (2010). Acoustic correlates of politeness: Prosodic and voice quality measures in polite and informal speech of Korean and German speakers. In *Proceedings of the fifth international conference on speech prosody*. Chicago.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Hawkins, S. (2010). Phonological features, auditory objects, and illusions. *Journal of Phonetics*, *38*, 60–89.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, *39*, 311–321.

Holtgraves, T. M. (2001). Politeness. In: W. P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 341–355). Chichester: John Wiley & Sons.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proceedings of the Institute of Electrical and Electronics Engineers international conference on acoustics, speech, and signal processing* (Vol. 1, pp. 669–672). Montreal.

Ito, M. (2004). Politeness and voice quality—The alternative method to measure aspiration noise. In *Proceedings of the second international conference on speech prosody* (pp. 213–216). Nara.

Irvine, J. T. (1979). Formality and informality in communicative events. *American Anthropologist*, *81*, 773–790.

Ismail, N., & Jemain, A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum* (pp. 103–158). Winter.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In: M. Lewis, & J. Haviland (Eds.), *Handbook of emotion* (pp. 220–235). New York: Guilford.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin*, *129*, 770–814.

Key, M. (1975). *Paralanguage and kinesics (Nonverbal communication)*. Metuchen, New Jersey: Scarecrow Press.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*, 97–120.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, *24*, 71–89.

Ng, S. K., & Reid, S. A. (2001). Power. In: W. P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 357–370). Chichester: John Wiley & Sons.

Ofuka, E. (1996). *Acoustic and perceptual analyses of politeness in Japanese speech*. Ph.D. Dissertation. University of Leeds.

Ofuka, E., McKeown, J. D., Waterman, M. G., & Roach, P. J. (2000). Prosodic cues for rated politeness in Japanese speech. *Speech Communication*, *32*, 199–217.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In: L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge: Cambridge University Press.

Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. In: A. Pavlenko, A. Brackledge, I. Piller, & M. Teutsch-Dwyer (Eds.), *Multilingualism, second language learning, and gender* (pp. 231–254). New York: Mouton de Gruyter.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America*, *109*, 1668–1680.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: The University of Michigan Press.

Podesva, R. J. (2011). Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics*, *39*, 233–264.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., & Bolker, B. (2012). *Generalized linear mixed models using AD Model Builder*. R package version 0.7.2.12.

Shin, S. (2005). *Grammaticalization of politeness: A contrastive study of German, English and Korean*. Ph.D. Dissertation. Berkeley: University of California.

Shue, Y. -L., Keating, P., Vicenik, C., & Yu, K. (2011). Voicesauce: A program for voice analysis. In *Proceedings of the 17th international congress of phonetic sciences*. Hong Kong.

Shrivastav, R., & Sapienza, C. (2003). Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America*, *114*, 2217–2224.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, *118*, 1677–1688.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*, 91–113.

Sohn, H.-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Talkin, D. (1995). A robust algorithm for pitch tracking (rapt). In: W. B. Kleijn, & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). New York: Elsevier.

Tatham, M., & Morton, K. (2004). *Expression in speech: Analysis and synthesis*. Oxford: Oxford University Press.

Trouvain, J. (1999). Phonological aspects of reading rate strategies. *Phonus*, *4*(Research Report Phonetics Saarbrücken), 15–35.

Trouvain, J., & Grice, M. (1999). The effect of tempo on prosodic structure. In *International proceedings of the 14th international congress of phonetic science* (pp. 1067–1070). San Francisco.

Trouvain, J. (2004). Tempo variation in speech production: Implications for speech synthesis. *Phonus*, 8 (Published Doctoral Dissertation).

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, *14*, 129–182.

Winter, B. (2011). Pseudoreplication in phonetic research. In *Proceedings of the international congress of phonetic science* (pp. 2137–2140). Hong Kong, August 2011.
Winter, B., & Christiansen, M. H. (2012). Robustness as a design feature of speech communication. In: T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *Proceedings of the 9th international conference on the evolution of language* (pp. 384–391). New Jersey: World Scientific.
Wish, M., Deutsch, M., & Kaplan, S. (1976). Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology*, *33*, 409–420.
Yoon, K.-J. (2004). Not just words: Korean social models and the use of honorifics. *Intercultural Pragmatics*, *1*, 189–210.
Yuan, C., & Li, A. (2007). The breath segment in expressive speech. *Computational Linguistics and Chinese Language Processing*, *12*, 17–32.
Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). New Jersey: Prentice Hall.