Original Article

Language *and* Speech

Loudness Trumps Pitch in Politeness Judgments: Evidence from Korean Deferential Speech

Language *and* Speech
1–26

DOI: 10.1177/0023830918824344
journals.sagepub.com/home/las
SAGE

Kaori Idemaru
University of Oregon, USA

Bodo Winter
University of Birmingham, UK

Lucien Brown
Monash University, Australia

Grace Eunhae Oh
Konkuk University, South Korea

Abstract

Social meaning is not conveyed through words alone, but also through how words are produced phonetically. This paper investigates the role of loudness and pitch in determining the perception of politeness-related judgments in Korean. It has been proposed that high pitch is universally associated with polite or deferential social meanings. In contrast to this, Experiment 1 examined the perceptual effect of pitch and found no effect. Experiment 2 tested the effect of loudness, and found that listeners associate quieter speech with deference. Finally, Experiment 3 investigated the simultaneous effects of loudness and pitch, and found again that loudness had a consistent effect, whereas pitch only had a weak effect. Analyses of individual differences suggest that in contrast to loudness, which is interpreted uniformly across Korean listeners, pitch has more variegated social meanings: Some listeners associate high pitch with deferential meaning, others associate low pitch with deferential meaning. Thus, we find loudness to be a more unambiguous indicator of deferential speech than pitch. These findings shed light on how different acoustic properties contribute to the indexing of social stances, and they suggest that the role of pitch in conveying politeness may have been overstated in past research.

Keywords

Sociophonetics, stance, politeness, perception, Korean

Corresponding author:
Kaori Idemaru, University of Oregon, 308 Friendly Hall, Eugene, OR 97403, USA.
Email: idemaru@uoregon.edu

Introduction

Speech is characterized by inherent variability. A range of factors influences the phonetic realization of any particular utterance. Such factors include the linguistic and social context, the speaker's gender, regional dialect, age, identity, speech style, and the speaker's mood and affective state, which change from moment to moment. These social and affective factors are transmitted by the acoustic speech signal in parallel with the linguistic content. The fact that speakers and listeners use acoustic features in systematic ways to transmit various social meanings (e.g., see Hay & Drager, 2007 for review) suggests very rich and complex representations of phonetic knowledge and processing. The social, emotional and expressive dimension of speech exists alongside the representational and referential functions. Any account of speech production and speech perception needs to consider these functions together.

The current study contributes to a growing body of research that examines how social stances related to politeness are conveyed by phonetic cues. Specifically, we focus on the perception of deferential and non-deferential utterances in Korean, examining the roles of fundamental frequency (F0) and intensity.[1] Korean is a suitable language for this kind of investigation since deferential and non-deferential stances are systematically differentiated through the use of grammaticized honorific forms. Every single utterance in Korean must obligatorily include honorific features, making the social deferential meanings we aim to investigate abundantly available in speech materials. Furthermore, the prior research on phonetic characterization of Korean politeness (Winter & Grawunder, 2012; Brown, Winter, Idemaru, & Grawunder, 2014) provides a foundation on which we can build this research to assess the relative contribution of different phonetic cues.

In this paper, we examine the proposal that high pitch is universally associated with politeness (see, e.g., Ohala, 1984; see also discussion in Winter & Grawunder, 2012). For our experiments, we acoustically manipulated F0 of spoken Korean utterances to determine whether the same utterance was perceived as more or less deferential if it was spoken in higher or lower pitch. To compare the influence of pitch, we also examined intensity (loudness) in another experiment. Our results show that loudness is a more consistent perceptual cue for deferential/non-deferential stances than pitch. An analysis of interindividual differences furthermore shows that there is great variability in how pitch is interpreted, with some listeners attributing deferential meaning to high pitched utterances and others attributing deferential meaning to low pitched utterances. In contrast to this, loudness is interpreted the same way by all participants. The fact that pitch is a more variable and less reliable cue for a deferential stance than loudness stands against the notion that there is a straightforward association between high pitch and politeness across languages.

1.1 Phonetic Investigation of Politeness-Related Meaning

Speakers produce speech sounds, many of which are linked to the social characteristics of the speakers (e.g., Labov, 1966). In addition, speakers may vary their speech from moment to moment. Some of this intraindividual variation functions as the performance of linguistic style, the construction of identity, and the expression of emotional and social stance (e.g., Hay, Jannedy, & Mendoza-Denton, 1999; Eckert, 2000; Bell, 2001; Podesva, 2007; Kiesling, 2009; Freeman, 2014). Our work focuses on the intraspeaker variation that indexes politeness and, more specifically, the social meanings of deferential and non-deferential stances in Korean.

The construct of politeness is a complex phenomenon that has been defined in many different ways (see Eelen, 2001, pp. 1–29; Culpeper, Haugh, & Kádár, 2017). Politeness manifests itself in numerous kinds of behaviors that are influenced by various communicative intentions (see for

example Kádár & Haugh, 2013). In this paper, we use the terms "politeness-related behaviors" and "politeness-related phenomena" to refer to an array of behaviors and phenomena that arise related to politeness. This study focuses on the context-appropriate adoption of deferential and non-deferential registers in Korean. In the pragmatics literature, this form of politeness has been referred to by terms such as "discernment politeness" (Ide, 1989) and "indexical politeness" (Brown, 2011). Following recent literature in pragmatics, we use the term "deference" as referring to "submitting to or showing regard to a superior or someone else deserving of respect" (Haugh, Chang, & Kadar, 2015, p. 81).

Within the pragmatics literature, prior research on deference has typically focused on lexical and grammatical characteristics. Recently, however, a body of research has started to emerge that supports the idea that phonetics is also important to the way that languages communicate politeness-related phenomena (see Brown & Prieto, 2017 for review). This research has shown that various features of speech covary with politeness contrasts, including F0 and intensity (Winter & Grawunder, 2012 for Korean; Ofuka, McKeown, Waterman, & Roach, 2000 for Japanese). Other features known to associate with polite language include slower speech rate (Ofuka et al., 2000 for Japanese; Lin, Kwock-Ping, & Fon, 2006 for Taiwanese; Gu, Zhang, & Fujisaki, 2011 for Mandarin; Santabalbina, 2013 for Spanish), increased breathiness (Campbell, 2004 for English; Ito, 2004 for Japanese; Winter & Grawunder, 2012 for Korean), and the use of high and sustained boundary tones or upstepped pitch accents (Orozco, 2010 for Mexican Spanish; Astruc, Vanrell, & Prieto, 2016 for Catalan).

1.2 Deferential and Non-deferential Registers in Korean

The tendency for pragmatics research to focus exclusively on verbal forms is particularly strong in the case of Korean. This is probably due to the fact that honorifics are such a salient characteristic of Korean grammar. Korean honorifics are a specialized grammatical system for marking deference and other related social meanings such as formality (see Lee & Ramsey, 2000, Chapter 7; Brown, 2011, Chapter 2). Although the system allows for the marking of various levels, the most basic distinction is between two general registers of speech, one being the honorific level known as *contaymal* ("respect speech") and the other being the non-honorific level known as *panmal* ("half-speech"). This simple binary contrast is the most basic and fundamental distinction within the Korean system of politeness. It is also a distinction to which Korean speakers are sensitive (Lee & Ramsey, 2000, p. 260). The two registers are distinguished by contrasting lexical items, as well as by honorific morphology on the verbs, in particular the addition of *–yo* and *–supnita* endings in *contaymal. Contaymal* is prototypically used when addressing elders, superiors, and adult strangers, whereas *panmal* is applied when interacting with intimate adults of equal or inferior age or rank, as well as with children (Lee & Ramsey, 2000, p. 251; Brown, 2011, p. 25). Thus, *contaymal* primarily indexes a deferential stance (i.e., showing regard to a superior), while *panmal* indexes a non-deferential stance. Although *contaymal* and *panmal* prototypically index these stances, this is not to say that there is a one-to-one correspondence. Indeed, the literature on Korean honorifics shows that *contaymal* may be used for other meanings and functions, such as marking psychological distance, formality, institutional identity and even sarcasm, whereas *panmal* may mark aggression, anger, or a submissive, childlike stance (see Brown 2015b). In this study, however, we focus on the deferential and non-deferential aspects of *contaymal* and *panmal* registers.

In addition to morphological marking, we now know that the phonetic characteristics of speech also robustly differentiate the two Korean registers, as shown by Winter and Grawunder (2012). Examining the speech of Korean speakers when addressing status superiors or status-equal friends, Winter and Grawunder (2012) presented a comprehensive "phonetic profile" of deferential and

non-deferential speech. In this laboratory study, Korean speakers produced unscripted utterances in a variety of simulated situations. In half of the situations, the imaginary interlocutor was a close friend or a younger acquaintance, thus eliciting non-deferential speech, and in the other half of the situations, the imaginary interlocutor was an elder stranger, a workplace superior, or a professor, eliciting deferential speech. Various F0 and intensity features (mean, SD, and range), voice quality features (local jitter, local shimmer, harmonics-to-noise ratio, and H1-H2), and speech rate features were measured. In addition, the number of different types of fillers were counted (oral fillers such as *oh*, nasal fillers such as *mh*, and breath intakes). The analysis revealed that almost all phonetic features and fillers analyzed in the study covaried with the distinction between deferential and non-deferential stances. The only exception was mean intensity, which showed no consistent differences between the two stances. Overall, the production study found that Korean deferential speech is lower-pitched, clearer (higher harmonics-to-noise-ratio, lower H1-H2), and less variable (smaller SDs and narrower range for F0 and intensity, smaller local jitter, smaller local shimmer). In addition, deferential speech was slower than non-deferential speech and included more audible breath intakes, more oral fillers, and fewer nasal fillers.

In a subsequent study, Brown et al. (2014) investigated whether phonetic features are perceptually relevant for detecting deferential and non-deferential stances in Korean. The study tested whether Korean listeners could use phonetic information alone to correctly perceive utterances intended for status superiors and utterances intended for status inferiors. The authors used phrases or clauses that were extracted from longer deferential and non-deferential utterances, so that these subparts did not include any lexical or grammatical honorific or non-honorific forms. The results indicated that, indeed, voice features do provide perceptual information for this distinction. Using only phonetic cues, Korean listeners could fairly accurately perceive deferential and non-deferential utterances (up to 70% correct).

Although Winter and Grawunder (2012) and Brown et al. (2014) demonstrated the importance of phonetic cues in the communication of social stance-taking, some important questions still remain unanswered. In particular, the stimuli in Brown et al. (2014) were characterized by the same natural covariation of acoustic features as was observed in Winter and Grawunder (2012). Due to the covariation of many potential acoustic cues, the perception experiment conducted by Brown et al. (2014) cannot be used to tease apart the relative contributions of different cues, and indeed, the experiment was not designed with this goal in mind. As such, while we now know that speaker's voice transmits deferential and non-deferential stances in Korean, we do not know *which* phonetic cues Korean listeners use in perceiving these meanings. To shed light on this question, we systematically manipulated acoustic features in the present series of experiments to investigate their effects in perception. We investigated the effect of F0 in Experiment 1 and the effect of intensity in Experiment 2. In Experiment 3, we investigated F0 and intensity together.

1.3 Perceptual Cues: F0 and Intensity

In the speech perception literature, there is a long history of empirical research demonstrating that phonetic categories are defined by multiple acoustic features. For example, as many as 16 acoustic features are reported to covary with the voicing distinction, such as between *bear* and *pear* (Lisker, 1986). Whereas any of these multiple features may inform category distinction, the perceptual role and effectiveness of each feature varies and is also subject to contextual factors. In the case of stop voicing at the onset position, English listeners make greater use of voice onset time than F0 (Abramson & Lisker, 1985; Idemaru & Holt, 2011) even though each reliably covaries with the voicing categories in production (Kingston & Diehl, 1994).

As a first foray into investigating perceptual phonetic cues to deferential and non-deferential speech in Korean, this study focuses on the independent influence of F0 and intensity. These two acoustic features were selected from a long list of covarying features (Winter & Grawunder 2012, Brown et al., 2014) for several reasons. F0 is arguably the most frequently studied acoustic feature related to politeness phenomena (e.g., Ofuka et al., 2000; Ohara, 2001 for Japanese; LaPlante & Ambady, 2003 for English; Nadeu & Prieto, 2011 for Catalan; Gu et al., 2011 for Mandarin). The importance of pitch was noted by Brown and Levinson (1987), a seminal publication on purported politeness universals. In their monograph, the authors included a short section on "Phonology and prosody" (p. 267–268), in which they predicted that a high-pitched voice should signal self-humbling stances and deference (p. 268). F0 has also been emphasized in many investigations of politeness-related phenomena due to Ohala's (1983, 1984) Frequency Code hypothesis. According to this "biological code" (see Gussenhoven, 2004: Ch. 5), a low-pitched voice signals dominance, and a high-pitched voice signals subdominance through the vocal projection of body size: Larger, more dominant animals emit lower-pitched sounds. There is considerable support for the idea that human and non-human animals interpret pitch with respect to dominance (e.g., Puts, Gaulin, & Verdolini, 2006; Puts, Hodges, Cárdenas, & Gaulin, 2007). However, it is an open question whether these dominance relations (often closely tied in with physical, rather than social dominance) are meaningfully related to politeness-related phenomena, and in particular whether high pitch is indeed universally perceived as conveying politeness-related meanings such as deference.

The literature on pitch in the context of politeness-related phenomena paints a mixed picture, with some studies finding high pitch to be associated with polite or deferential meanings, and others finding the same for low pitch. For Japanese (Ofuka et al., 2000; Ohara, 2001) and American English (LaPlante & Ambady, 2003), researchers found that high-pitched utterances were judged to be more polite. On the other hand, Nadeu and Prieto (2011) found that Catalan speakers relied more on the speaker's facial expression than F0 of the voice when these two types of information were both available in the stimuli. Our production results from Korean (Winter & Grawunder, 2012) contradict Ohala's and Brown and Levinson's view of high pitch more directly: We found that Korean speakers spoke in *lower* F0 when speaking in a deferential register (see Shin, 2005 for an earlier report on Korean; see Irvine, 1979 for an early anecdotal report for the language Mursi). In a similar vein, Hübscher, Borràs-Comes, and Prieto (2017) found that pitch was lowered in Catalan utterances that were spoken in a formal, deferential speech style. The authors also found slower speech rate, lower intensity, and more stable and clearer voice in formal speech, and interpreted these patterns as a form of "prosodic mitigation," a strategy designed to soften aspects of speech (Caffi, 2007).

Given the production data (Winter & Grawunder, 2012), we predicted that lower F0 is perceived as signaling deference and higher F0 is perceived as signaling non-deference in Korean, the opposite to Ohala's proposal. If our prediction is supported, Ohala's (1984) characterization of politeness and associated pitch does not explain the pattern observed for Korean deferential speech very well. Such results would suggest that a different underlying meaning, other than subdominance, may exist to connect low pitched voice and deference.

The potential association between lower F0 and deference (and potentially other politeness-related social meanings) may also contradict another biological principle, the Effort Code, proposed by Gussenhoven (2002) (see also Gussenhoven, 2004: Ch. 5). The Effort Code proposes that there is a relationship between sounds and meaning through the effort and energy expenditure required of sound vocalization. The Effort Code predicts that wider F0 excursions and raised F0, reflecting more effort on the part of the speaker, are associated with meanings such as being emphatic, significant, surprised, and obliging (Gussenhoven, 2002). Addressing a status superior

may involve actively signaling that the speaker requires more effort to produce an utterance; or it may involve actually expending more effort in planning one's utterance (see also Winter & Grawunder, 2012). Both of these factors would predict that wider F0 excursions and raised F0 should be associated with deferential speech, contrary to what we and others (Shin, 2005) have found for Korean so far.

We chose intensity as another phonetic feature to investigate alongside F0. Intensity has received relatively little attention to date in phonetic studies of social meaning in general, and of politeness in particular. Besides F0, a lot of work has examined the social meanings of voice quality (Campbell, 2004 for English; Ito, 2004 for Japanese; Winter & Grawunder, 2012 for Korean) and speech rate (Ofuka et al., 2000 for Japanese; Lin, et al., 2006 for Taiwanese; Santabalbina, 2013 for Spanish; Gu et al., 2011 for Mandarin). There are, however, several reasons to investigate the influence of intensity in this study. On a practical level, F0 and intensity are easy to manipulate systematically and such manipulation has been widely used in phonetic research. In contrast, manipulation and synthesis of voice quality (i.e., breathy, modal, and creaky voicing) is still at an early stage (Brikholz, Kröger, & Neuschaefer-Rube, 2011; Xu, Lee, Wu, Liu, & Birkholz, 2013). More importantly, on a theoretical level, if politeness is conveyed through the projection of larger body size (dominance) versus smaller body size (subdominance) as proposed by Ohala (1984), it is possible that intensity is also used as a size code, since larger animals generally produce louder sounds. By analogy to the proposal for pitch, a quieter voice may communicate subdominance and, therefore, also be associated with politeness-related stances such as deference. Although the Effort Code (Gussenhoven 2002), just like the Frequency Code, focuses on the relationship between intonational patterns and social meanings, its principles may be extended to include other features such as intensity. It would predict that a louder voice, requiring more energy expenditure, communicates a deferential stance. The role of intensity is also interesting from the perspective of the proposal by Hübscher et al. (2017) that signaling deference is a matter of prosodic mitigation. Such strategy involves, among other things, speaking with a quieter voice.

Within our own research on Korean, the role of intensity is still unclear. Winter and Grawunder (2012) reported only a numerical trend toward deferential speech being quieter (not statistically reliable). In contrast, the stimuli of Brown et al. (2014) showed robust covariance of intensity with deferential speech. Although correlation between production cues and perception cues is often reported (e.g., Newman, 2003; Perkell et al., 2004) and sometimes assumed for reasons intrinsic to theory (e.g., Fowler, 1986), there is evidence for the lack of such correspondence in other cases (Abramson & Lisker 1985; Idemaru, Holt, & Seltman, 2012; Idemaru & Holt, 2013). A direct test of the perceptual role of intensity in the Korean deferential stance distinction should clarify this question.

1.4 The Role of Gender

The previous literature suggests that the effect of phonetic cues in the interpretation of deference may be differentially influenced by the gender of interlocutors. For example, Ohara (2001) and Loveday (1981) reported more robust use of pitch in the deferential speech of Japanese women, as opposed to Japanese men. In the case of Korean, Winter and Grawunder (2012) found that men and women were largely similar in their expression of politeness. However, the differences between deferential and non-deferential speech were more pronounced for some acoustic variables for female speakers, compared to male speakers. Whereas both genders showed clear differences for all F0 features, voice quality features, and speech rate, only female speakers showed clear differences for intensity variability, intensity range, and speech rate. Given these production results, we explored the role of speaker and listener gender in interpreting deferent and intimate relationships of the

interlocutors. The role of gender is also important to study because past research on Korean has suggested that the differential use of honorifics may have a gendered dimension (Brown, 2015a).

1.5 The Current Study

Our prior production findings (Winter & Grawunder, 2012; Brown et al., 2014) led us to develop the following hypotheses:

1. Two acoustic features F0 and intensity covary with deferential and non-deferential registers in Korean. Given this, we hypothesized that the two features influence perception such that low F0 utterances relative to high F0 utterances as well as quiet utterances relative to loud utterances are judged as speech spoken in a deferential context.
2. Between the two acoustic features, F0 has been shown to display the more robust and consistent covariation with the two registers. Given this, we hypothesized that F0 would show a more robust influence on perception.

To test these hypotheses, three perceptual experiments were conducted. In all experiments, a subset of the speech materials from Brown et al. (2014) were used as stimuli. The stimuli did not contain any morphological markings indicating deferential or non-deferential stance (i.e., no honorifics). In each experiment, the stimuli were systematically manipulated so that mean F0 varied in five levels in Experiment 1, mean intensity varied in five levels in Experiment 2, and both mean F0 and intensity independently varied in three levels in Experiment 3. Half of the stimuli were spoken by female voices, half by male voices. Similarly, half of the listeners were female, and half were male. Native Korean listeners heard the stimuli and judged whether they had been spoken to a status superior or a status inferior. Listeners' responses were analyzed to examine to what extent F0 (Experiment 1) and intensity (Experiment 2) influenced auditory perception of deferential and non-deferential meaning, and to examine how F0 and intensity together affected auditory perception (Experiment 3). We did not have a specific hypothesis regarding the gender of stimulus speaker and listeners; however, the influence of speaker and listener gender was explored in each experiment. The results of all three experiments are discussed together in the general Discussion section.

2 Experiment 1: Effect of Pitch

In this experiment, we investigated the effect of F0 in the perception of deferential and non-deferential speech. Korean listeners heard utterances that systematically varied in F0, and judged whether they were spoken to a status superior or a status inferior.

2.1 Methods

2.1.1 Participants. Twenty-four adult native Korean listeners (12 female and 12 male) participated for a small payment. At the time of testing, they were all students at a university in Seoul, and on average 23.3 years old (range 20–27).[2] All participants were recruited and tested in Seoul, South Korea. Thirteen participants were born and raised in the Seoul/Gyeonggi area; the other 11 originated from other areas. Because all dialects of Korean have a grammaticized deferential/non-deferential distinction (and because standardized deferential speech is heavily promulgated across all dialect areas through the education system; see King, 2006), we did not expect any dialectal difference in the perception of the two registers. The data from all 24 listeners were included in the analysis. All participants reported to have normal hearing.

2.1.2 Speech materials. Korean speech samples from Brown et al. (2014) served as the baseline stimuli ("original"). The original speech samples were productions of predetermined utterances issuing requests in 10 different contexts ("scenarios"). Eight out of the 10 scenarios were randomly selected for this experiment to keep the total number of stimuli manageable. The scenarios are listed in the Appendix.

Eight native Korean speakers (four female and four male), all speakers of Standard Seoul Korean, produced the utterances in two social situations: In one, they simulated speaking to an unfamiliar professor, eliciting deferential speech, and in the other, they simulated speaking to their best friend, eliciting non-deferential speech. Speakers were recorded individually, seated in front of a computer in a sound-attenuated booth. The stimulus items were presented visually in Korean orthography on a computer screen via E-Prime (Schneider, Eschman, & Zuccolotto, 2002). Whereas the 10 scenarios were presented randomly, the deferential version and the non-deferential version within each scenario were presented sequentially. The intention behind this methodological choice was to elicit more differentiation of the two versions by placing them back to back. Whether the deferential version or the non-deferential version was presented first was randomized across scenarios. We also displayed a picture of the imagined interlocutor on the computer screen at each trial so that the speaker could more easily imagine an actual speaking situation. With the deferential stimuli, we displayed a picture of a man that contained a number of semiotic cues for authority, elderliness, and scholarliness (suit and tie, glasses, gray hair, fountain pen, books in the background, etc.). With the non-deferential stimuli, we displayed a digital image of a close friend of the participant, which each participant supplied before the recording session.

Each speaker read each item twice. Speakers were permitted further productions if they were not satisfied with their performance or if they produced disfluencies. In selecting from the productions, our general policy was to use the final production. In cases where the final production included disfluencies, we used the second-to-last production. The productions were recorded using a Marantz PMD670 solid-state recorder and a Shure SM10A head-mounted microphone at a sampling rate of 44.1 kHz and 16-bit quantization. The microphone was placed approximately 1–1.5 inch away from the speaker's mouth. The recording level remained the same across speakers.[3] An experimenter stayed in the booth with the participant to operate E-Prime and ensure proper functioning of the recording equipment.

We used the deferential version of the scenarios as the base stimuli in this study. The deferential version of one of the eight scenarios is shown here as an example:

1. 교수님, 지난번에 말씀하신 컴퓨터 프로그램을 구했습니다. 그런데 사용법이 어려워서 그러는데 바쁘시겠지만 잠깐만 가르쳐 주실 수 있나요?
2. kyoswu-**nim**, cinan pen-ey **malssum**ha-**si**-n khemphyuthe, phulokulaym-ul kwuha-yss-**supnita**, kulentey sayongpep-i elyew-ese kule-nuntey, pappu-**si**-kyess-ciman camkkan-man kaluchy-e cwu-**si**-l swu iss-na-**yo**?
3. "Professor, I've bought that computer program you mentioned last time. But the instructions are difficult. I know you must be busy, but can you teach me how to use it?"

As is typical with deferential utterances in Korean, this example sentence contains various grammatical and lexical honorific markers, indicated by the bold segments in (2). The suffix *–nim* renders address terms honorific, *malssum* is an honorific noun form referring to "words" (of a status superior), *-si-* is the subject honorific verb ending, and *–supnita* and *–yo* are deferential endings. However, notice that the underlined middle clause does not include any such markers. Taken by itself, there is no lexical or morphological information in this clause that indicates the intended

Table 1. The mean F0 (Hz) of stimuli in each F0 Manipulation level.

	−16%	−8%	Original	8%	16%	Mean
Female 1	166	183	197	213	228	197
Female 2	158	172	187	202	217	187
Female 3	187	205	223	241	259	223
Female 4	161	177	192	207	223	192
Female mean	168	184	199	216	232	200
Male 1	79	87	93	100	108	93
Male 2	84	92	99	108	115	100
Male 3	95	104	112	121	131	113
Male 4	102	112	122	131	141	122
Male mean	90	99	107	115	124	107

F0: fundamental frequency; Hz: Hertz.

register, rendering it morphologically ambiguous with respect to whether it indexes a deferential stance or not.

Using the Pitch Synchronous Overlap Add Method function in Praat (Boersma and Weenink, 2010), and its "multiply pitch frequencies" function specifically, we manipulated the F0 of the baseline stimuli so that the mean F0 was raised and lowered by 8% and 16%. The overall contour was preserved. This resulted in five versions of each clause: +16%, +8%, the original, −8%, and −16%. All stimuli were then amplitude-normalized to 70 dB to control for intensity variation. In the resulting set of stimuli, the mean female F0 varied from 168 Hz to 232 Hz across five F0 manipulation levels with a mean of 200 Hz, and the mean male F0 varied from 90 Hz to 124 Hz with a mean of 107 Hz (Table 1).

Our choice of 8% and 16% as manipulation levels stems from the fact that in previous production studies, the difference between deferential and non-deferential speech was 17 Hz (Winter & Grawunder, 2012) and 16 Hz (Brown et al., 2014). Changing F0 upward and downward by 8% approximates this empirically established production difference. Changing by 16% allowed us to explore exaggerated levels that go beyond what was observed in the production data. The F0 manipulation resulted in 320 stimuli (8 speakers × 8 scenarios × 5 pitch levels).

2.1.3 Procedure. Before the start of the experiment, participants filled out a language background questionnaire and received instruction regarding the task in Korean. They were instructed that they would hear short phrases in Korean and that their task was to judge whether each utterance was spoken to "someone above the speaker" or "someone below the speaker." These are culturally meaningful descriptions that are known to associate closely with the use of deferential and non-deferential speech, respectively (Park, 2007; Yoon, 2004). We also used this way of operationalizing deference successfully in Brown et al. (2014). Following our earlier work, we interpret listeners' response of "someone above the speaker" as the indication that the listeners judged the utterance to be deferential (i.e., showing regard to a superior). Similarly, we interpret listeners' response of "someone below the speaker" as the indication that the listeners judged the utterance to be non-deferential. Listeners were specifically instructed that the utterances were phrases devoid of honorific markings, and that judgments would need to be made based on the sound or feeling of the utterance.

Seated in front of a computer monitor, participants heard each stimulus once through headphones and then judged whether the utterance was spoken to "someone above the speaker" or

"someone below the speaker." Participants responded by pressing the number-8 key or the number-2 key on the number pad. The number-8 key on the number pad (the "up" arrow key) was assigned for "above the speaker" and the number-2 key (the "down" arrow key) was assigned for "below the speaker." These keys were selected due to the metaphorical relationship between verticality and power (Schubert, 2005; Giessner & Schubert, 2007) and because we used the same key alignment in Brown et al. (2014). The response categories written in Korean and the corresponding numbers for the keys appeared on screen at each trial.

The experiment was delivered by E-Prime experiment software (Schneider, et al., 2002). Participants had 3 s to respond to each trial, and after each response was registered, there was a 1-s interval before the next trial began. The 320 stimuli were blocked by speakers (i.e., participants heard all 40 of the utterances produced by one speaker, before moving on to the next speaker). The reason for blocking the stimuli in this way was to allow participants to tune into the subtle individual characteristics of the utterances produced by each speaker. Mixing different speakers would have exposed participants to acoustic variability besides that connected to deferential speech (see Brown et al., 2014). The order of speaker blocks was randomized. In addition, all stimuli were randomized within each speaker block. Participants had a chance to take a break after each speaker block. All participants completed the task in approximately 30 min.

2.1.4 Analysis. We used R and the packages "lme4" 1.1.15 (Bates, Maechler, Bolder, & Walker, 2015b), "afex" 0.16.1 (Singmann, Bolker, Westfall, & Aust, 2016), and "RePsychLing" 0.0.4 (Baayen, Bates, Kliegl, & Vasishth, 2015) for the linear mixed effects model analysis reported below. The package "MuMIn" 1.15.6 (Bartoń, 2016) was used to compute R^2 values for mixed models (Experiment 3) following the approach outlined in Nakagawa and Schielzeth (2013). For data processing, we used the packages "tidyverse" 1.1.1 (Wickham, 2016b) and "stringr" 1.2.0 (Wickham, 2016a). We used linear mixed effects models because of our repeated measures design to account for the fact that we have multiple non-independent data points per listener, per speaking voice and per item (scenario). We used mixed logistic regression because our dependent variable was binary, that is, "someone above" versus "someone below" decisions. The fixed effects predictors were F0 Manipulation (five levels), Log Reaction Time, Trial Order, Speaker Gender (gender of the stimulus voices), and Listener Gender. F0 Manipulation is our critical test variable; all other variables are control variables. The inclusion of Log Reaction Time was motivated by the finding of Brown et al. (2014) that politeness judgments differed based on response speed. The inclusion of Trial Order was motivated because of the potential for participants changing their overall levels of "someone above" or "someone below" responses. For the two factors Speaker Gender and Listener Gender, we also considered their interaction, as it is plausible that women/men would react differently to voices from their own gender, as opposed to voices from another gender. The F0 Manipulation variable was coded as a continuous factor (results did not change if it was coded as a categorical or ordinal factor).

In the random effects component, we included random intercepts for Listener, Speaker, and Item (cf., Baayen, Davidson, & Bates, 2008). In addition, we initially fitted by-Listener, by-Speaker and by-Item random slopes for our critical variable of interest, namely F0 Manipulation (compare Barr, Levy, Scheepers, & Tily, 2013). However, a look at the random effects structure using Principal Components Analysis (see Bates, Kliegl, Vasishth, & Baayen, 2015a) revealed that the by-Speaker and by-Item random slopes were not supported by the data (see also Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Hence, we included only the by-Listener slopes in the final model. In previous experiments using the same stimuli (without acoustic manipulation), we also found little differences between the different items or the different speaking voices. Including by-Listener slopes in the final model is furthermore necessary to investigate inter-individual variation

Table 2. Percentage of "someone above the speaker" (deferential) responses as a function of F0 Manipulation.

	% "someone above"
−16%	46%
−8%	52%
Original	49%
+8%	50%
+16%	50%

F0: fundamental frequency.

in how listeners respond to our perceptual manipulation (see below). The following formula (in lme4 syntax) shows the structure of our main model:[4]

4. Response ~ F0 Manipulation + Log Reaction Time + (Trial Order) + Speaker Gender * Listener Gender + (1 + F0 Manipulation|Listener) + (1|Speaker) + (1|Item)

The Trial Order variable led to convergence issues in estimating the model and was excluded from the main comparisons of the other effects, since it is a control variable anyway. Exclusion of this variable did not change the interpretation of the main result. All categorical and continuous variables participating in interactions were sum-coded or centered to facilitate their interpretation. All *p*-values are based on likelihood ratio tests of the model with the effect in question against the model without the effect in question. In line with standards of reproducible research, the scripts and data are made available with this publication and can be retrieved on the following publicly accessible repository: https://github.com/bodowinter/korean_pitch_intensity/.

2.2 Results

To our surprise, the results showed that there was *no* statistically reliable main effect of F0 Manipulation, $\chi^2(1) = 0.04$, $p = 0.85$. A look at the model's coefficients revealed a positive relationship between F0 and "someone above the speaker" (deferential) judgments (logit estimate = 0.014, $SE = 0.075$), however, this was not statistically reliable and numerically negligible (a one-step increase along the F0 continuum changed the odds of observing a "someone above the speaker" response by only 1.01 to 1). As shown in Table 2, descriptive averages reveal no consistent response pattern across the five conditions.

A look at the by-subject varying slopes for the effect of F0 Manipulation revealed a potential reason for why the overall effect may not have been statistically reliable. Whereas 14 listeners were estimated to have positive slopes for the effect of F0 Manipulation (interpreting utterances with higher pitch as spoken to "someone above"), 10 listeners were estimated to have negative slopes (interpreting utterances with lower pitch as spoken to "someone above"). Thus, our sample of listeners was split roughly in half as to how they interpreted the social meaning of pitch with regard to deferential stance.

A number of supporting variables resulted in statistically reliable differences. There were effects of Log Reaction Time, $\chi^2(1) = 113.37$, $p < 0.0001$, and Trial Order, $\chi^2(1) = 15.54$, $p < 0.0001$. Slower responses were relatively more likely to be "someone above" responses (logit estimate = 0.22, $SE = 0.02$). This suggests that when participants respond in a more deliberate fashion, they may have a bias towards the deferential response option. In addition, the effect of Trial Order showed

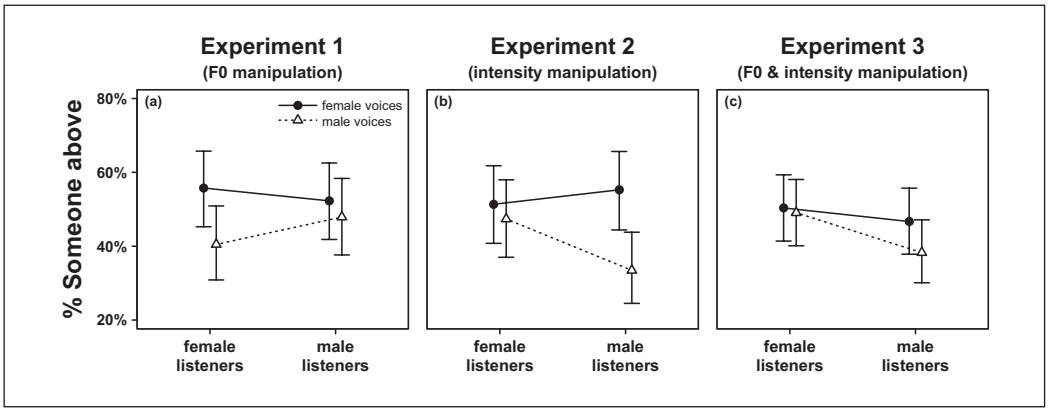

Figure 1. Gender interaction effects for Experiment 1 (a), Experiment 2 (b), and Experiment 3 (c): points indicate predictions of the mixed effect model with 95% confidence intervals.

that over the course of the experiment, participants became more likely to choose the "someone above" response option (logit estimate, per trial = 0.001, $SE = 0.0003$). Thus, over the time course of the experiment, participants were more likely to default to the deferential response option.

We also found gender effects. There was a statistically reliable interaction between Speaker Gender and Listener Gender, $\chi^2(1) = 20.39$, $p < 0.0001$; logit estimate: 0.44, $SE = 0.98$. There also was a marginally reliable main effect of Speaker Gender, $\chi^2(1) = 3.74$, $p = 0.052$; logit estimate: −0.40, $SE = 0.18$. There was no main effect of Listener Gender, $\chi^2(1) = 0.34$, $p = 0.56$; logit estimate: 0.08, $SE = 0.14$. The obtained gender effects are most easily understood by looking at Figure 1(a). As can be seen, male voice stimuli overall led to fewer "someone above" responses (main effect), and particularly so with female listeners (interaction effect). However, it was not the case that male or female listeners had an overall difference in how often they selected "someone above" response (no reliable Listener Gender main effect).

3 Experiment 2: Effect of Intensity

In this experiment, we investigated the effect of intensity in the perception of deferential and non-deferential speech. Korean listeners heard utterances that systematically varied in intensity, and judged whether they were spoken to a status superior or a status inferior.

3.1 Methods

3.1.1 Participants. Twenty-three adult native Korean listeners (12 female, 11 male) participated for a small payment. None of these listeners participated in Experiment 1. At the time of testing, they were all students at a university in Seoul, and on average 23.6 years old (range 19−28). All participants were recruited and tested in Seoul, Korea. Twelve were born and raised in the Seoul/Gyeonggi area; the other 11 originated from other areas. All participants reported normal hearing.

3.1.2 Speech materials. The same set of original speech samples from Brown et al. (2014) used in the Experiment 1 served as the baseline stimuli. We manipulated the intensity of the baseline stimuli using the "scale intensity" function on Praat so that the mean intensity was lowered by

Table 3. The mean intensity (dB) of stimuli in each Intensity Manipulation level.

	−10%	−7.50%	−5%	−2.50%	Original	Mean
Female 1	54	55	57	59	60	57
Female 2	66	68	70	72	73	70
Female 3	62	64	65	67	69	65
Female 4	67	68	70	72	74	70
Female mean	62	64	66	67	69	66
Male 1	60	62	63	65	67	63
Male 2	62	63	65	67	68	65
Male 3	65	67	69	71	73	69
Male 4	64	66	67	69	71	67
Male mean	63	64	66	68	70	66

2.5%, 5%, 7.5%, and 10%. This method was adopted because some of the baseline original samples had pressure values close to 1.0, and thus increasing the intensity of the samples would result in clipping these sounds. The first author listened to all stimuli to make sure the speech was clearly audible even when the intensity was reduced by 10%. This resulted in five versions of each clause: 0% dB change ("original"), −2.5%, −5%, −7%, and −10%. In the resulting set of stimuli, the mean female intensity varied from 62 dB to 69 dB across five Intensity Manipulation levels with the mean of 66 dB, and the mean male intensity varied from 63 dB to 70 dB with the mean of 66 dB (Table 3). In our previous production study, the intensity difference between deferential and non-deferential speech conditions was 1.3 dB on average, and 3.5 dB by the speaker who made the largest difference (Brown et al., 2014). The change of intensity by 2.5% upward and downward from −5% level (the middle level among the five) approximates this difference. Changing by 5% from −5% level approximates the maximum difference we observed in the production. The overall mean F0 was then normalized to 200 Hz in female stimuli, and to 100 Hz in male stimuli to control for F0 variation. The intensity manipulation resulted in 320 stimuli (8 speakers × 8 scenarios × 5 intensity levels).

3.1.3 Procedure and analysis. The procedure was exactly the same as Experiment 1. The statistical analysis was also exactly the same, except that Intensity Manipulation assumed the role of F0 Manipulation (also including the same random slopes, etc.). As before, we treated this main condition variable as a continuous effect, with each step-size indicating a change by one of the condition levels.

3.2 Results

Our results revealed a statistically reliable main effect of Intensity Manipulation, $\chi^2(1) = 15.00$, $p = 0.0001$, logit estimate: −0.12, $SE = 0.03$. As shown in Table 4, descriptive averages revealed a consistent response pattern across the five conditions, with "someone above the speaker" (deferential) responses being less likely with relatively louder stimuli. Stimuli with intensity lowered by 10% resulted in the largest proportion of deferential responses, 52%. The original stimulus (which was 10% louder) was judged to be deferential only 41% of the time.

In terms of individual differences, the picture was much different from the pitch results obtained in Experiment 1. This time, a look at the estimated by-subject slopes of Intensity Manipulation revealed that *all* listeners interpreted lower intensity to indicate deferential speech (with "someone

Table 4. Percentage of "someone above" (deferential) responses as a function of Intensity Manipulation.

	% "someone above"
Original	41%
−2.5%	45%
−5.0%	48%
−7.5%	51%
−10%	52%

above" responses). This means that in contrast to the F0 Manipulation of the previous experiment, there were no differing interpretations as to the meaning of intensity as an acoustic cue.

Again, a number of supporting variables also resulted in statistically reliable differences. There were effects of Log Reaction Time, $\chi^2(1) = 91.76$, $p < 0.0001$, logit: 0.21, $SE = 0.02$, and Trial Order, $\chi^2(1) = 70.93$, $p < 0.0001$, logit: 0.002, $SE = 0.0003$. As in Experiment 1, slower responses were relatively more likely to be "someone above" responses and over the course of the experiment, participants became more likely to choose the "someone above" response.

We also found another Speaker Gender and Listener Gender interaction, $\chi^2(1) = 54.83$, $p < 0.0001$, logit: −0.74, $SE = 0.1$, as well as a statistically reliable effect of Speaker Gender, $\chi^2(1) = 6.02$, $p = 0.01$, logit: −0.53, $SE = 0.17$, but not of Listener Gender, $\chi^2(1) = 1.01$, $p = 0.32$, logit: −0.21, $SE = 0.21$. These results are displayed in Figure 1(b). Again, male speakers (male voice stimuli) overall led to fewer "someone above" responses (main effect). This time, however, the gender interaction was in the opposite direction from Experiment 1: The male listeners were particularly likely to judge that female stimuli were deferential more frequently than male stimuli.

4 Experiment 3: Effects of F0 and Intensity Together

In this experiment, we investigated the role of F0 and intensity in the perception of deferential and non-deferential speech. Korean listeners heard utterances that simultaneously varied in F0 and intensity. As in previous experiments, participants were asked to judge whether each utterance was spoken to a status superior or a status inferior.

4.1 Methods

4.1.1 Participants. Sixteen adult native Korean listeners (eight female and eight male) participated for a small payment. None of these listeners participated in Experiments 1 or 2. At the time of testing, they were all students at a university in Seoul, and on average 23.3 years old (range 19–28). All participants were recruited and tested in Seoul, Korea. Fifteen of the 16 participants were born and raised in the Seoul/Gyeonggi area, with one participant born in Gyeongsang province. All participants reported normal hearing.

4.1.2 Speech materials. A subset of the original speech samples from Brown et al. (2014) used in Experiments 1 and 2 served as the baseline stimuli. The subset consisted of the speech produced by four speakers (two male and two female). We specifically chose the four speakers whose speech was most accurately identified as deferential/non-deferential in Brown et al. (2014). The target features F0 and intensity were manipulated so that they varied independently in three levels in the stimuli.

Table 5. The mean F0 (Hz) and intensity (dB) of stimuli in each Manipulation level.

	F0			Intensity		
	High	Mid	Low	High	Mid	Low
Female 1	226	195	164	58	55	52
Female 3	261	225	189	69	66	62
Male 1	113	97	81	68	65	61
Male 2	121	104	87	68	64	61

F0: fundamental frequency.

We controlled the variation of F0 and intensity more uniformly in this experiment than in Experiments 1 and 2. Recall that, for example, the F0 of the original utterances was raised and lowered by 8% and 16% in Experiment 1. The stimuli, therefore, contained natural by-speaker variation of F0 (Table 1), and in addition, by-item ("scenario") variation of F0 even within speaker. For example, the mean F0 of the +8% condition for Female 1 stimuli ranged from 201 Hz to 231 Hz across items. To eliminate this by-item variation in the target features, F0 and intensity were manipulated so that the resulting utterances had consistent mean Hz and dB values for all items ("scenarios") within each manipulation level (i.e., high, low, and mid), but we allowed these target mean values to vary across speakers. The target Hz and dB values for each of the three factor levels are reported in Table 5. This means that, for example, in the case of female speaker 1, all eight items in the high F0 condition had the exact mean F0 of 226 Hz. The target manipulation values of F0 were derived as the grand mean for each speaker across all items (Mid), 16% raised value of the grand mean (High), and 16% lowered value (Low). The target manipulation values of intensity were derived as the grand mean for each speaker across all items (High), 5% lowered value (Mid), and 10% lowered value (Low). The three levels of F0 and the three levels of intensity were paired for all combinations. This resulted in a total of 288 stimuli (4 speakers × 8 scenarios × 3 F0 manipulation levels × 3 intensity manipulation levels).

4.1.3 Procedure and analysis. The procedure was exactly the same as Experiments 1 and 2. In this analysis, the condition predictors were F0 Manipulation (three levels, entered as continuous predictors) and Intensity Manipulation (three levels), as well as their interaction. The supporting variables were the same as in the previous experiments. Since both F0 Manipulation and Intensity Manipulation are critical variables of interest, we included by-Speaker and by-Listener random slopes for both of these fixed effects.

4.2 Results

Our results revealed a statistically reliable main effect of Intensity Manipulation, $\chi^2(1) = 10.38$, $p = 0.001$, logit: −0.27, $SE = 0.07$, as well as a barely statistically reliable effect of F0 Manipulation, $\chi^2(1) = 4.24$, $p = 0.04$, logit: 0.37, $SE = 0.17$. There was no statistically reliable interaction between F0 Manipulation and Intensity Manipulation, $\chi^2(1) = 2.05$, $p = 0.15$, logit estimate: 0.07, $SE = 0.05$. As shown in Table 6, descriptive averages suggest a numerically large difference between low F0 (39% "someone above" responses) and high F0 stimuli (53% "someone above" responses); however, as we will discuss below, there was also large individual variability. For intensity, the results were consistent with the findings from Experiment 2: Quieter stimuli were most likely responded to as "spoken to someone above" (52%), compared to louder stimuli (41%). Taken together, a higher pitched and quieter voice was more likely to be judged as deferential.

Table 6. Percentage of "someone above" (deferential) responses as a function of F0 Manipulation and Intensity Manipulation.

	% "someone above": F0	% "someone above": Intensity
Low	39%	52%
Mid	49%	47%
High	53%	41%

F0: fundamental frequency.

How did the weight of pitch and loudness compare in listeners' deferential/non-deferential judgments? In terms of descriptive percentages, the effects of the two acoustic manipulations were comparable. In fact, of the two effects, F0 Manipulation had a steeper slope (log odd estimate 0.37 as opposed to −0.27 for Intensity Manipulation). However, a look at by-listener variation reveals that again, *all* listeners interpreted loudness the same way (all Intensity Manipulation slopes were negative). In contrast to this, five listeners had negative slopes for F0 Manipulation, interpreting high pitch as being non-deferential ("someone below") and 11 listeners had positive slopes, thus interpreting high pitch as being deferential ("someone above"). This is also reflected in the estimated random slope SDs for listeners. These SDs quantify the amount of by-listener variation, and they are only 0.24 for the Intensity Manipulation slopes, in contrast to 0.66 for the F0 Manipulation slopes. Comparing R^2 across models with and without Intensity Manipulation or F0 Manipulation random slopes reveals that a model with by-listener-varying slopes for F0 Manipulation describes 7.6% of the variance in deferential/non-deferential judgments (conditional R^2; see Nakagawa & Schielzeth, 2013). In contrast, the by-listener-varying slopes for Intensity Manipulation describe only 1.4% of the variance. These results reaffirm that Intensity Manipulation had much more *consistent* effects across listeners than F0 Manipulation.

Again, there was an effect of Log Reaction Time, $\chi^2(1) = 67.85$, $p < 0.0001$, logit: 0.23, $SE = 0.03$, but this time no effect of Trial Order, $\chi^2(1) = 0.04$, $p = 0.88$, logit: −0.000094, $SE = 0.0004$. As in Experiments 1 and 2, slower responses were relatively more likely to be "someone above" responses (logit estimate = 0.23, $SE = 0.03$).

We also found another Speaker Gender and Listener Gender interaction, $\chi^2(1) = 5.34$, $p = 0.02$, logit: −0.29, $SE = 0.13$. This time, there were no Speaker Gender, $\chi^2(1) = 1.58$, $p = 0.21$, logit: −0.2, $SE = 0.14$, and Listener Gender, $\chi^2(1) = 2.85$, $p = 0.09$, logit: −0.29, $SE = 0.16$, main effects. These results are displayed in Figure 1(c). The gender interaction followed a similar pattern as observed in Experiment 2 (and not as in Experiment 1). Again, it was the male listeners in particular who were much less likely to judge that male stimuli were deferential (speaking to "someone above"). However, this time, female listeners also had a slight tendency to perceive female voices as being less likely to be deferential.

To shed light on the differing signs of the Speaker Gender and Listener Gender interactions across experiments, the data from all experiments was pooled in a meta-analysis using one mixed model on all responses (with random intercepts for Listener, Speaker and Item). There was a reliable Speaker Gender and Listener Gender interaction effect, $\chi^2(1) = 6.44$, $p = 0.01$, logit: −0.15, $SE = 0.06$, and a reliable Speaker Gender main effect, $\chi^2(1) = 5.28$, $p = 0.02$, logit: −0.38, $SE = 0.14$, but no Listener Gender main effect, $\chi^2(1) = 0.36$, $p = 0.55$, logit: −0.07, $SE = 0.11$. The main effect of Speaker Gender has the following directionality: Across all experiments, there was a tendency for all listeners (regardless of their gender) to indicate that a male voice was more likely to address "someone below the speaker" (descriptive mean: 52%) compared to "someone above the speaker" (44%). The interaction effect modulated this pattern: Male listeners were more likely

to indicate that male voices were addressed to "someone below the speaker" (non-deferential) (53%) than "someone above the speaker" (42%). The responses of female listeners patterned the same way but somewhat less extremely so, with 45% responding "someone above the speaker" when listening to male voices and 52% when listening to female voices.

5 General Discussion

This study investigated the role of pitch (F0) and loudness (intensity) in the auditory perception of deferential and non-deferential social meanings in Korean speech. Assuming that Korean listeners are sensitive to the covariation pattern of acoustic properties in the speech signal and determine perceptual cues based on their informativeness (i.e., predictive power), we hypothesized that Korean listeners would rely on both F0 and intensity in making deferential and non-deferential judgements (Hypothesis 1), giving more perceptual weight to F0 relative to intensity (Hypothesis 2). However, neither hypothesis was borne out in the current study. Contrary to our hypothesis, we found that Korean listeners respond to intensity more consistently in identifying deferential and non-deferential utterances: Quieter utterances are perceived as deferential and louder utterances are perceived as non-deferential. In contrast to the consistent response pattern to intensity, we found large individual differences in how Korean listeners use F0 in identifying the social meanings: Some listeners associate high pitch with a deferential utterance, whereas others associate low pitch with a deferential utterance.

The null result for the effect of F0 in Experiment 1 does not necessarily indicate the absence of an overall group effect. In fact, we did find a barely statistically reliable effect in Experiment 3, suggesting that deferential speech was more frequently linked to higher F0 than to lower F0. However, these group-mean results must be interpreted carefully, given the presence of the noteworthy individual differences that are revealed by investigating the predicted by-listener varying random slopes of the mixed effects model analysis. Across Experiments 1 and 3, 25 listeners associated high F0 with deferential stance and 15 listeners had the reverse association. Thus, there is not a particularly strong preference across listeners to associate high pitch (or low pitch) with deferential meanings. In contrast, across Experiments 2 and 3, 39 listeners associated low intensity with deferential speech and 0 listeners had the reverse mapping. The comparison between the F0 and intensity results highlights the critical finding that listeners used intensity more *consistently*, with a uniform mapping between phonetics and social meaning. The direction of the mapping for intensity in perception also matched the pattern we found in production. In contrast, the finding that more speakers associated higher F0 with deference is actually the opposite of production findings: Deferential speech consistently has lower F0 (Winter & Grawunder, 2012). Although F0 does appear to play a role in perception, its effect is more variable across different listeners.

We note the relatively small magnitude of the effect of our manipulations: Listeners' politeness judgments varied only between averages of 41% and 52% in our experiments. However, this does not detract from the key finding that pitch and loudness influence perception of deferential and non-deferential styles in Korean, with loudness having a more systematic influence across listeners. The small magnitude of the effects for pitch and intensity furthermore suggests that other phonetic cues, not investigated in this study, may play a role. This will be discussed below.

The current study has important implications for the understanding of the role of pitch in politeness-related phenomena. Pitch has often been considered one of the key phonetic features, if not the most important phonetic feature, that signals politeness cross-linguistically (Ohala, 1984; Loveday, 1981; Ohara, 2001; LaPlante & Ambady, 2003). The current results call into question this generalization and suggest that the role of pitch in politeness-related phenomena may vary cross-linguistically. In Korean, at least between pitch (F0) and loudness (intensity), loudness is consistent

(all listeners use it) and uniform (all listeners use it in the same direction) as a perceptual cue for deference. Since the mapping between pitch and deference varies across listeners as we saw in the data, the status of pitch as a perceptual cue to deferential meaning is questionable for Korean.

These results contradict Ohala's (1984) proposal that high pitch is cross-linguistically associated with politeness-related phenomena such as deference (Frequency Code). However, they are not inconsistent with the fundamental idea from which the Frequency Code was derived. At the base of the Frequency Code hypothesis is the idea that social meanings associated with pitch are mediated by body size projection: low pitch, which projects a larger body size, communicates aggression, assertiveness, dominance, etcetera.; and high pitch, which projects a smaller body size, conveys social subordinacy, politeness, a non-threatening stance, etcetera. (Ohala, 1984). The interpretation of the Frequency Code as a "Size Code" has indeed been proposed by Gussenhoven (2002: 47). If projected body size mediates such social meanings, it is possible that these meanings are associated not only with pitch but also with intensity variation. A larger body is likely to have a larger respiratory system, and thus an association can be made between larger body size and a louder voice. Perhaps projected body size underlies the transmission and interpretation of deferential and non-deferential stances in Korean, as Ohala (1984) hypothesized, but body size may also be projected by the loudness rather than the pitch of the voice alone.

Our intensity results must also be interpreted cautiously vis-a-vis the Effort Code (Gussenhoven, 2002). All else being equal, the Effort Code would associate a louder voice with deference. However, we did not control for spectral slope, which more robustly signals vocal effort (e.g., Nordenberg & Sundberg, 2004). Thus, the relationship between intensity and the Effort Code needs to be addressed in future research with careful control of features correlating vocal efforts, such as sound pressure levels and spectral slope.

The current results have shown that Korean speakers may listen for loudness/quietness in a voice in deciphering whether a deferential stance is being conveyed. Recall that in Korean speech production, F0 is consistently lower in deferential utterances than in non-deferential utterances. Here, we speculate that one of Korean speakers' acoustic and auditory targets, when intending to speak deferentially, may be lowered intensity, and that low F0 might simply be an articulatory consequence of lowering intensity (e.g., Titze & Sundberg, 1992). To speak quietly, the speaker typically decreases the rate of airflow, which in turn limits the rate of vocal fold vibration resulting in lower F0. This view is consistent with the proposal by Hübscher et al. (2017) that "prosodic mitigation" is what matters in the signaling of deference: A mitigated voice is a quieter voice and may, as a result of decreased loudness, also have decreased F0. Winter and Grawunder (2012) furthermore discuss the literature on affective and emotional vocal displays, in which high pitch is generally associated with animatedness (as well as anger and happiness), which may interfere with the intention of speakers to "mitigate" their speech. The current results are also consistent with some of the findings of Rilliard, Erickson, De Moraes, and Shochi (2014), particularly about a specific type of politeness expression in Japanese conveying the mitigation of negative face-threatening acts: This type of expression also showed mitigated prosody.

Thus, the relationship between F0 and deferential speech seems either potentially epiphenomenal (mediated through loudness) or potentially affected by other affective meanings of pitch, such as animatedness, in Korean. In the absence of a norm for interpreting F0 directly with respect to deferential/non-deferential meanings, Korean listeners perhaps assign different social meanings to F0, depending on their interpretation of the task. Some listeners may indeed interpret high F0 as indexing submissiveness (e.g., Ohala, 1984), or perhaps as signaling friendliness or "positive politeness" (e.g., Brown & Levinson, 1987). Others, however, may interpret it as being too animated and lively to be addressed to a superior (cf. discussion in Winter & Grawunder, 2012; Hübscher et al., 2017). In addition, as discussed in Mason, Domínguez, Winter, and Grignolio

(2015), F0 expresses not only social attitudes and emotions, but also grammatical information, and even semantic characteristics of what is being talked about (e.g., Perlman, Clark, & Johansson Falck, 2014). In the case of Korean, F0 additionally expresses lexical information on some consonants (Silva, 2006). Mason et al. (2015) discuss how communicative signals, such as increased versus decreased F0, are inherently "pluripotential": precisely because F0 serves so many different functions in language (including social, emotional, grammatical, semantic functions), different listeners may allocate different meanings to high or low pitch.

The findings in this study, in particular the different patterns of listener responses to F0 and intensity, underscore the importance of considering the phonetics of politeness-related meanings from a holistic perspective, where the intended social meanings covary with numerous different acoustic factors (Winter & Grawunder, 2012; Hübscher et al., 2017), not just a few isolated cues. When indexing politeness-related meanings, speakers are faced with the complex task of upholding the face of the interlocutor while at the same time projecting a socially appropriate or desired representation of themselves (see Spencer-Oatey & Ruhi, 2007). Under such pressure, especially in a cultural context where politeness-related behaviors are of such pivotal importance as is the case in Korean society (see Harkness 2015, pp. 493–495 on honorifics as "an emblem of Korean sociality"), it is reasonable that speakers may exploit multiple redundant features to ensure successful communication (Hailman, 2008; Winter, 2014). A multiplicity of cues that listeners can draw from is especially important in the case of phenomena that vary across speakers and listeners (cf. Winter 2014). In the case of Korean, for example, if one cue is interpreted differently by some listeners (such as pitch), another cue (such as loudness) could help in signaling the intended deferential meaning. Thus, signaling social meanings via an array of acoustic features may help to counteract the pluripotentiality of any given cue, such as in this case, F0.

In fact, our results do imply the presence of multiple perceptual cues in the Korean deferential/non-deferential distinction. In Brown et al. (2014), Korean listeners identified deferential and non-deferential utterances with 70% accuracy based solely on phonetic features. Since the stimuli in Brown et al. (2014) were clauses extracted from deferential and non-deferential productions, they retained all phonetic cues intact. In the current study, in which only F0 and intensity were allowed to vary, the listeners' identification function varied only between 41% and 52%. These results appear to suggest that other important cues were likely missing from the stimuli. As discussed in the Introduction, production studies in various languages, including Korean, have found that polite utterances have a higher degree of breathiness and a slower speech rate (Ofuka et al., 2000; Lin et al., 2006; Winter & Grawunder, 2012; Santabalbina, 2013). These features may serve as perceptual cues in Korean. Even though the synthesis of voice quality is still at an early stage, it has recently been used in speech studies (e.g., Xu et al., 2013). Future research employing such technology will shed light on the perceptual role of voice quality in perceiving deferential speech.

It is also important to examine whether there may be specific locations in an utterance that attract concentrations of cues. Relevant to this point are the findings of Ofuka et al. (2000) on Japanese. The study found that in addition to speech rate (analyzed as a global feature of an utterance), the duration and F0 movement of the final vowel (a localized feature) had critical effects on the perception of deferential speech in Japanese. In fact, the duration and F0 characteristics of the final vowel "changed the total impression of the utterance" (p. 214) with regard to its deferential meaning. In verb-final agglutinative languages such as Japanese and Korean, utterance final position is rich in pragmatic and discourse meanings in general, and the use of different prosodic boundary tones have been shown to assign a variety of different stances (see Park, 2012). Together, this indicates that the cues to social stances may be localized at specific points in an utterance. In the current study, the acoustic cues were manipulated across the entirety of the stimulus utterances.

Future studies are needed to explore whether there are "hot spots" for the acoustic cues that signal social meanings in Korean. Manipulation of acoustic features specific to such hot spots, rather than overall manipulation, may in fact elicit more robust differences between deferential versus non-deferential responses.

It should of course be recognized that our results pertain to a certain mode of politeness, namely the distinction between deferential and non-deferential speech. Even within this distinction, our data deals with a limited range of social situations. The stimuli used in the current study involved the speech act of making a request (and this is true also for Hübscher et al., 2017, and for most of the stimuli in Winter & Grawunder, 2012). As such, it is possible that these studies have tapped into a specific mode of deference that includes the expression of embarrassment when imposing a burden on a status superior (see Rilliard et al., 2017; Rilliard, Shochi, Martin, Erickson, & Aubergé, 2009). Future research needs to determine whether the current findings hold also for other politeness attitudes in Korean, given the complex range of politeness subcategories that may be available (see Rilliard et al., 2017, 2009). The results leave open the question of whether the Frequency Code may explain other kinds of politeness-related attitudes in Korean. However, we can say that at least for the mode of politeness investigated here, the role of the Frequency Code may have been overstated.

The current results also revealed interesting gender effects. Male voices were overall more likely to be perceived as speaking to a social inferior ("someone below") compared to female voices. Across male and female stimuli, male F0s were overall lower than female F0s, with no overlap between the two (see Table 1). Although experiment trials were blocked for speaker, it is entirely possible that listeners responded to F0 variation across blocks, that is, across speakers and their gender. Thus, the effect of gender translates so that higher F0 (spoken by females) was more likely to be perceived as deferential. This is consistent with the results of Experiment 3, and also with the prediction of the Frequency Code. However, we would like to interpret this evidence first and foremost as a genuine gender effect, rather than a result that is in line with the Frequency Code. First of all, the Frequency Code makes predictions for within-speaker variation in F0 (see Ohala, 1984). The fact that we found no effect of F0 manipulation but an effect of gender (which co-varies with F0) suggests to us that a cultural explanation is more likely, especially since past research has reported on the gendered meanings of Korean politeness (Brown, 2015a). In particular, we think that our results can be seen as consistent with the cross-cultural phenomena whereby "at a stereotypical level, politeness is often considered to be a woman's concern" (Mills, 2003, p. 203). Men, on the other hand, may have more social leeway to instigate casual or impolite language and behaviors (see Holmes, 1995). These assumptions may exist in South Korean society, where traditional patriarchal systems are still prevalent (Cho, 2006, p. 190). Patriarchy is known to influence Korean language usage in various ways, including other areas of politeness and indexicality such as the use of address terms (see Kim, 2015).

Due to the cultural expectation that men assume higher positions in society in general and speak less politely, listeners may simply be less likely to associate male speech as being spoken to a social superior. Thus, a Frequency Code-based explanation does not mesh as well with what we know about politeness in Korean society. Furthermore, the fact that it was male listeners in particular who displayed this tendency suggests that it is men who may hold these expectations more strongly.[5] Explaining the gender effect in terms of the Frequency Code (due to the natural co-variation in pitch) does not easily explain the interaction we found between the listener's and the speaker's gender. Nonetheless, it has to be stated that the perception of male voices (with lower voice pitch) as less deferential is at least consistent with an expression of the Frequency Code across speaking voices, although the Frequency Code would fail to explain the within-speaker variation of our acoustic manipulation.

6 Conclusion

The current study presented a sociophonetic investigation of politeness-related phenomena. We have examined the perceptual cues to Korean deferential versus non-deferential speech registers. In Korean, the social meaning of deference is codified systematically and extensively in its grammar and lexicon through verbal honorific forms, yet this study, together with Winter and Grawunder (2012) and Brown et al. (2014), demonstrates that phonetics has an important role in transmitting the social meanings. We believe that our findings underscore the importance of phonetic investigation in politeness-related phenomena.

Various phonetic features are known to covary across Korean deferential and non-deferential speech in production (Winter & Grawunder 2012). The current study verified the role of intensity in perceiving the two stances, whereas the systematic perceptual role of F0 was not verified. Based on these findings, we argue that between intensity and F0, intensity is a more robust cue for the deferential/non-deferential speech distinction in Korean: Quieter utterances signal deference, and louder utterances signal non-deference. The current findings emphasize the importance of perceptual verification of acoustic cues: even a strongly informative acoustic cue (F0 in this case) may not necessarily have a robust and consistent role in perception. The variation of listeners' responses to F0 in the current study suggest that Korean listeners may assign different social and affective meanings to F0 such as submissiveness, friendliness, animatedness and so forth, which would have different consequences for deference/non-deference judgments.

The findings also represent a case study in which an influential theory of the relationship between F0 and social meanings, Ohala's Frequency Code, does not account for the data very well. The results suggest that there may be context-, culture-, or language-specific interpretations of what it means to be polite. The phonetics of politeness clearly matters, but which precise acoustic cues are used differ by context, language, and between listeners.

Acknowledgements

A portion of this work was presented at the 18th International Congress of Phonetic Sciences, Glasgow, Scotland. August 14–15, 2015. We thank Sara King and Sue Yoon for assistance with the experiments.

Funding

This work was supported by an Academy of Korean Studies grant (AKS-2015-R24).

Notes

1. Throughout the paper, acoustics terms (i.e., "F0" and "intensity") are used to refer to phonetic features when the reference is primarily acoustic, whereas when it is primarily auditory, auditory terms (i.e., "pitch" and "loudness") are used.
2. In the current study, we tested a total of 64 Korean listeners with ages ranging from 19 to 28. Future studies may want to address potential age effects on phonetic cues to politeness.
3. At the recording, we did not conduct intensity calibration, and the original sound pressure levels were not retained. Our focus here, however, is not articulatory events, but rather the manipulation of the auditory effect of loudness.
4. In a preliminary analysis, we explored whether there were any quadratic effects of F0 Manipulation (following the finding of U-shaped response curves in politeness phenomena, Ofuka et al., 2000; Idemaru, Winter, & Brown, 2015). However, these analyses showed no evidence for non-linearity and hence F0 Manipulation was entered into the model as a simple linear continuous factor. We also initially tested for any interactions between our main test variable, F0 Manipulation, and the Gender variables. Not finding any statistically reliable effects, we proceeded to fit F0 Manipulation as a simple main effect, without

interactions. In addition to these analyses, we tested the critical effects of interest (F0 Manipulation and in Experiment 2, Intensity Manipulation) when these effects were coded as categorical factors or as ordered factors—the substantive interpretation of the results reported in the main body in the text does not change when these alternative models are considered. Decisions about the appropriate random effects structure for this design were made on the basis of Experiment 1 and this random effects structure was carried over to the other experiments.

5. It should be noted, however, that such expectations do not necessarily reflect language production. At least in some spheres, previous studies suggest that it is Korean men who may actually use deferential speech more frequently (see Brown, 2015a).

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). Orlando, FL: Academic Press.

Astruc, L., Vanrell, M. M., & Prieto, P. (2016). Cost of the action and social distance affect the selection of question intonation in Catalan. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Intonational grammar in Ibero-romance: Approaches across linguistic subfields* (pp. 91–114). Amsterdam, the Netherlands: John Benjamins.

Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). RePsychLing: Data sets from psychology and linguistics experiments. R package version 0.0.4. Retrieved from https://github.com/dmbates/RePsychLing

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bartoń, K. (2016). MuMIn: Multi-Model Inference. R package version 1.15.6 [Computer software]. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *arXiv*:1506.04967. Retrieved from https://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge, UK: Cambridge University Press.

Birkholz, P., Kröger, B. J., & Neuschaefer-Rube, C. (2011). Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis. In *INTERSPEECH 2011: 12th Annual Conference of the International Speech Communication Association (ISCA)*, 28–31 August 2011, Florence, Italy (pp. 2681–2684). Retrieved from https://pdfs.semanticscholar.org/bedc/755c08b58765c0e84c1594dfa5a0737d6379.pdf

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Version 5.1. 35, Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Brown, L. (2011). *Korean honorifics and politeness in second language learning*. Amsterdam, the Netherlands: John Benjamins.

Brown, L. (2015a). Expressive, social and gendered meanings of Korean honorifics. *Korean Linguistics*, *17*(2), 243–267.

Brown, L. (2015b). Honorifics and politeness. In L. Brown & J. Yeon (Eds.). *The handbook of Korean linguistics* (pp. 303–309). Malden, MA: Wiley-Blackwell.

Brown, L., & Prieto, P. (2017). (Im)politeness: Prosody and gesture. In J. Culpeper, M. Haugh, & D. Z. Kádár (Eds.), *The palgrave handbook of linguistic (im)politeness* (pp. 357–379). London, UK: Palgrave Macmillan.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, *66*, 45–60.

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

Caffi, C. (2007). *Mitigation*. Oxford, UK: Elsevier.
Campbell, N. (2004). Accounting for voice-quality variation. In *Proceedings of the second International Conference on Speech Prosody* (pp. 217–220), *Nara, Japan*.
Cho, Y. (2006). Gender differences in Korean speech. In H. Sohn (Ed.), *Korean language in culture and society* (pp. 189–198). Honolulu, HI: University of Hawai'i Press.
Culpeper, J., Haugh, M., & Kádár, D. Z. (2017). *The palgrave handbook of linguistic (im)politeness*. London, UK: Palgrave Macmillan.
Eckert, P. (2000). *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. Malden, MA: Blackwell.
Eelen, G. (2001). *A critique of politeness theories*. Manchester, UK: St. Jerome Publishing.
Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. In I. G. Mattingly & N. O'Brien (Eds.), *Status report on speech research* (pp. 139–169). New Haven, CT: Haskins Laboratories.
Freeman, V. (2014). Hyperarticulation as a signal of stance. *Journal of Phonetics*, *45*, 1–11.
Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes*, *104*(1), 30–44.
Gu, W., Zhang, T., & Fujisaki, H. (2011). Prosodic analysis and perception of Mandarin utterances conveying attitudes. In *INTERSPEECH 2011: 12th Annual Conference of the International Speech Communication Association (ISCA)*, 28–31 August 2011, Florence, Italy (pp. 1069–1072).
Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In B. Bel & I. Marlien (Eds.), *Proceedings of speech prosody* (pp. 47–57). Aix-en Provence, France: Université de Provence.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
Hailman, J. P. (2008). *Coding and redundancy: Man-made and animal-evolved signals*. Cambridge, MA: Harvard University Press.
Harkness, N. (2015). Linguistic emblems of South Korean society. In L. Brown & J. Yeon (Eds.). *The handbook of Korean linguistics* (pp. 492–508). Malden, MA: Wiley-Blackwell.
Haugh, M., Chang, W. L. M., & Kadar, D. Z. (2015). Doing deference: identities and relational practices in Chinese online discussion boards. *Pragmatics: Quarterly Publication of the International Pragmatics Association*, *25*(1), 73–98.
Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, *36*, 89–103.
Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical frequency, referee design and style. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1389–1392). Berkeley, CA: University of California.
Holmes, J. (1995). *Men, women and politeness*. London, UK: Longman.
Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017). Prosodic attenuation characterizes formal register in Catalan requests. *Journal of Phonetics*, *65*, 145–159.
Ide, S. (1989). Formal forms and discernment: Two neglected aspects of universals of linguistic politeness. *Multilingua*, *8*(2–3), 223–248.
Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956.
Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English/r/-/l/. *The Journal of the Acoustical Society of America*, *133*(6), 4232–4246.
Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950–3964.
Idemaru, J., Winter, B., & Brown, L. (2015). The role of pitch in perceiving politeness in Korean. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.
Irvine, J. T. (1979). Formality and informality in communicative events. *American Anthropologist*, *81*(4), 773–790.
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, Japan*.
Kádár, D. Z., & Haugh, M. (2013). *Understanding politeness*. Cambridge, UK: Cambridge University Press.

Kiesling, S. F. (2009). Style as stance: Stances as the explanation for patterns of sociolinguistic variation. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives* (pp. 171–194). Oxford, UK: Oxford University Press.
Kim, M. (2015). Women's talk, mothers' work: Korean mothers' address terms, solidarity, and power. *Discourse Studies*, *17*(5), 551–582.
King, R. (2006). Dialectal variation in Korean. In H. Sohn (Ed.), *Korean language in culture and society* (pp. 264–280). Honolulu, HI: University of Hawai'i Press.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454.
Labov, W. (1966). Hypercorrection by the lower middle class as a factor in linguistic change. In *Sociolinguistics*. The Hague, the Netherlands: Mouton, 84, 102.
LaPlante, D. A., & Ambady, N. (2003). On how things are said: Voice tone, voice intensity, verbal content, and perceptions of politeness. *Journal of Language and Social Psychology*, *22*(4), 434–441.
Lee, I., & Ramsey, S. R. (2000). *The Korean language*. New York, NY: State University of New York Press.
Lin, H., Kwock-Ping, J.T., & Fon, J. (2006). An acoustic study on the paralinguistic prosody in the politeness talk in Taiwan Mandarin. *Proceedings from ISCA Tutorial and Research Workshop on Experimental Linguistics* (pp. 173–176). Athens, Greece.
Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and Speech*, *29*(1), 3–11.
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*(1), 71–89.
Mason, P., Domínguez, D. J. F., Winter, B., & Grignolio, A. (2015). Hidden in plain view: Degeneracy in complex systems. *BioSystems*, *128*, 1–8.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
Mills, S. (2003). *Gender and politeness* (Vol. 17). Cambridge, UK: Cambridge University Press.
Nadeu, M., & Prieto, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics*, *43*(3), 841–854.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.
Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America*, *113*, 2850–2860.
Nordenberg, M., & Sundberg, J. (2004). Effect on LTAS of vocal loudness variation. *Logopedics Phoniatrics Vocology*, *29*(4), 183–191.
Ofuka, E., McKeown, J. D., Waterman, M., & Roach, P. (2000). Prosodic cues for rated politeness in Japanese speech. *Speech Communication*, *32*, 199–217.
Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*(1), 1–18.
Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.
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, NY: Mouton de Gruyter.
Orozco, L. (2010). Estudio sociolingüístico de la cortesía en tratamientos y peticiones. Datos de Guadalajara. PhD thesis, El Colegio de México, Mexico.
Park, M. J. (2012). *The meaning of Korean prosodic boundary tones*. Leiden, the Netherlands: Brill.
Park, Y.-S. (2007). *Korean pragmatics*. Seoul, Korea: Pagijong Press.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts, *Journal of the Acoustical Society of America*, *116*, 2338–2344.
Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic prosody in story reading. *Cognitive Science*, *39*, 1348–1368.
Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.
Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.
Rilliard, A., Erickson, D., De Moraes, J. A., & Shochi, T. (2014). Cross-cultural perception of some Japanese politeness and impoliteness expressions. *Linguistic Approaches to Emotions in Context*, 251–276.
Rilliard, A., Shochi, T., Martin, J. C., Erickson, D., & Aubergé, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech*, *52*(2–3), 223–243.
Santabalbina, M. R. (2013). *Duración y percepción de la cortesía en español (experimento piloto)* [Duration and perception of the courtesy in Spanish (pilot study)]. Lugar: Valencia, 411.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh, PA: Psychology Software Tools.
Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1.
Shin, S. (2005). Grammaticalization of politeness: A contrastive study of German, English and Korean. Ph.D. Dissertation. Berkeley, CA: University of California.
Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*, 287–308.
Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. Retrieved from https://CRAN.R-project.org/package=afex.
Spencer-Oatey, H., & Ruhi, Ş. (2007). Identity, face and (im)politeness. *Journal of Pragmatics*, *39*(4), 635–638.
Titze, I. R., & Sundberg, J. (1992). Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America*, *91*(5), 2936–2946.
Wickham, H. (2016a). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.1.0. Retrieved from https://CRAN.R-project.org/package=stringr.
Wickham, H. (2016b). tidyverse: Easily Install and Load "Tidyverse" Packages. R package version 1.0.0. Retrieved from http://project.org/package=tidyverse.
Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, *36*, 960–967.
Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, *40*(6), 808–815.
Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PloS ONE*, *8*(4), e62397.
Yoon, K. J. (2004). Not just words: Korean social models and the use of honorifics. *Intercultural Pragmatics*, *1*(2), 189–210.

Appendix

Eight deferential scenarios used for the experiments. Underlined clauses were extracted and used as base stimuli.

1. 교수님, 지난 번에 말씀하신 책 있잖아요. 그 책이 도서관에 없는 거 같은데 빌려 주실 수 있을까요?
   Do you know the book you mentioned last time? It seems like they don't have the book in the library; can I borrow yours?
2. 교수님, 오늘 오후에 교수님을 뵙기로 했잖아요. 갑자기 급한 일이 생겨서 그러는데 오늘 대신 내일 뵐 수 있을까요?
   We were going to meet this afternoon (in your office). But something urgent has come up; can we meet tomorrow instead?

3. 교수님, 지난 주에 모든 학생들에게 이메일을 보내셨다고 하셨잖아요. 그런데 그 이메일을 못 받은 것 같은데 죄송하지만 다시 한 번 보내 주실 수 있나요?
   You said that you sent an e-mail to all students/friends last week. But I don't think I received that e-mail; can you send it again?

4. 교수님, 이번 주말에 엠티 가잖아요. 그런데 엠티 장소를 정확히 몰라서 그러는데 가는 길 좀 알려 주세요.
   We are having MT this weekend. But I don't know the MT location exactly; can you tell me where it is?

5. 교수님, 지난 수업에 "파라메트릭"이라는 말을 많이 사용하셨잖아요. 그 단어가 무슨 말인지 잘 모르겠는데 다시 한 번만 더 설명해 주세요?
   Last class/study group meeting you used the word "pragmatics" a lot. I don't really know what this word means; can you explain it again?

6. 교수님, 지난 번에 말씀하신 웹사이트 있잖아요. 그런데 그 사이트를 못 찾아서 그러는데 그 웹사이트 URL 한 번 더 가르쳐 주시겠어요?
   You know that website you mentioned last time. But I can't find that site; can you let me know the URL one more time?

7. 교수님, 지난번에 말씀하신 컴퓨터 프로그램을 구했습니다. 그런데 사용법이 어려워서 그러는데 바쁘시겠지만 잠깐만 가르쳐 주실 수 있나요?
   I've bought that computer program you mentioned last time. But the instructions are difficult; can you teach me how to use it?

8. 교수님, 제가 도서관에 가서 그 책을 가져다 드릴게요. 그런데 가방이 좀 무거워서 그러는데 여기에 가방 좀 놓고 갔다 와도 될까요?
   I will go to the library and fetch that book. But my bag is a bit heavy; can I leave my bag here?