Prosodic cues for rated politeness in Japanese speech

Etsuko Ofuka a,*, J. Denis McKeown b, Mitch G. Waterman b, Peter J. Roach c

a Department of Foreign Languages and Bilingual Studies, Eastern Michigan University, Ypsilanti MI 48197, USA
b School of Psychology, University of Leeds, Leeds LS2 9JT, UK
c Department of Linguistic Science, University of Reading, Whiteknights, Reading, RG6 6AA, UK

Received 16 October 1998; received in revised form 18 August 1999; accepted 22 November 1999

Abstract

In order to examine potential acoustic cues for politeness in Japanese speech, F0 and temporal aspects of polite and casual utterances of two question sentences spoken by 6 male native speakers were acoustically analysed. The analysis showed that F0 movement of the final part of utterances and speech rate of utterance were consistently differently used in these different speaking styles across all the speakers. Perceptual experiments with listeners confirmed that these acoustic variables, which were manipulated using digital resynthesis, had an impact on politeness perception. It was shown that how the final intonation of a sentence is spoken had a great impact on politeness judgements. In some cases the duration and F0 characteristics of the final vowel did change the overall impression of the utterance’s politeness. An experiment which used speech rate variations of a polite utterance showed the important role of this variable in perceived politeness. Politeness ratings showed an inverted U-shape as a function of speech rate, but differed according to particular speakers. The speech rate of listeners was found to affect their utterance rate preference; listeners preferred rates close to their own. These findings suggest that listener characteristics should be considered important in politeness speech research. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Prosody; Politeness; Japanese speech; Cue manipulation

1. Introduction

In any language community, the speakers’ ability to show an appropriate level of politeness is very important for smooth social interaction. It is especially so for the Japanese speech community, because Japanese society still attaches much importance to the hierarchical social relationship, which has been described as a ‘vertical society’ (Nakane, 1967, 1970) depending on various factors such as age, sex and social status; the appropriate use of politeness acknowledges and maintains this social hierarchy (Matsumoto, 1988). In the Japanese language, the level of politeness is encoded in a special linguistic system called keigo (literally meaning ‘terms of respect’), having special expressions or words for displaying respect and modesty towards addressees and persons referred to by the speaker. Japanese speakers select linguistic forms in the system, together with appropriate non-verbal forms such as body and facial expressions, tone of voice and appearance, according to the required level of politeness.

Since linguistic forms, attitudes and appearance have been regarded as very important in conveying politeness, these are the subjects of various kinds...
of textbooks and formal teaching at school and in
the work place in Japan. However, little attention
has been given to teaching how to make utterances
sound polite, though tone of voice is considered to
be important by native speakers (Ogino and Hong,
1992). Recently, the rapid increase in the number
of households and persons owning telephones and
the importance of telephones especially in business
have been changing attitudes towards speech: the
importance of speech and therefore the importance
of learning how to speak are beginning to be rec-
ognised. Nippon Housou Kyoukai (the Japanese
state broadcasting institute) has recently started to
broadcast a series of educational programmes on
the radio which focus on various aspects of spoken
language including the right usage of keigo, good
pronunciation and speaking styles and good
manners for telephone conversations (NHK,
1995).

The present study focuses on speech in relation
to politeness perception. In fact, the importance of
learning how to speak politely has been recognised
by native speakers. Ogino and Hong (1992) con-
ducted a questionnaire survey on what cues
Japanese people would use to evaluate the level of
politeness of the speaker with more than 200
Tokyo residents between 23 and 74 years of age. This
survey showed that a Japanese person would
mostly rely on the appropriateness of the speaker’s
use of keigo, followed by facial expressions, tone
of voice, gaze, gesture and clothes or shoes. Hong
(1992) also mentioned the importance of research
on the way of speaking in terms of conveying
politeness from the point of view of teaching
Japanese as a foreign language. He presented
about 100 native speakers of Japanese with polite
utterances spoken by 6 native speakers and the
same sentences spoken by 6 learners of Japanese
and asked them to judge whether the utterances
sounded polite or not to them. The results showed
that the polite utterances spoken by the learners
were not perceived as polite by more than half of
the native listeners, while the native utterances
were appropriately identified by more than 80% of
listeners. He concluded that it was probably due to
the incorrect prosody imposed on the utterances
by the learners. So if we can discover what makes
utterances polite or non-polite, it will be very
useful in terms of foreign language education and
it will also contribute to making synthetic speech
sound more natural.

The present psychophysical study focused on
the impression of politeness in relation to variables
at the acoustic level. The following approach was
adopted to investigate the acoustic cues to Japa-
nese politeness. This approach consists of three
stages:
1. recording polite and non-polite utterances
   (Section 2);
2. conducting a preliminary investigation of
   acoustic variables which appear to be relevant
to the degree of politeness conveyed in the
   recorded utterances (Section 3);
3. conducting perceptual experiments with stimuli
   which were created by computer cue manipula-
tion of potentially relevant acoustic variables,
   using selected utterances from stage 1 as source
   utterances, in order to observe the effects of
   these manipulated variables on politeness judg-
   ments (Sections 4 and 5).

2. Recording

There are two different ways of collecting
speech samples (Murray and Arnott, 1993); one is
‘field’ recordings in natural settings and the other
is laboratory recordings. There is, of course, a
trade-off here between realism on the one hand
and achieving a high degree of control on the
other. In fact, Scherer (1979a, p. 509) writes that
“simulated portrayals ... may be the only realistic
possibility for the research”. As has been noted by
a number of researchers, asking subjects to speak
test passages in a polite or angry way, e.g. often
induces theatrical exaggeration (Cosmides, 1983).
We therefore adopted a role-play method in which
speakers are given scenarios and asked to play
their roles (Williams and Stevens, 1972; Scherer
and Scherer, 1980; Cosmides, 1983; Hong, 1993).
Since the form of text used has been found to be
significant in signalling personality attributions or
emotions (Apple et al., 1979; Scherer et al., 1984),
there are at least three ways to eliminate or mini-
mise the effects of the content. One approach is
simply to use meaningless content, such as letters
of the alphabet (Davitz and Davitz, 1959; Green and Cliff, 1975) or nonsense syllables or isolated words (Levin and Lord, 1975). Another approach is to use semantically neutral phrases (Fairbanks and Pronovost, 1939). Finally, verbal content itself may be eliminated by using low-pass filtered speech materials (Starkweather, 1956; Scherer et al. 1984).

In the present study, we used test sentences with semantically neutral content. Since politeness is usually closely associated with appropriateness in a specific situation, it is difficult to separate it from verbal content and therefore the content would be an indispensable part of the judgement. The sentences used are:

(i) **Nimotsu-wa koredake desuka**, meaning ‘Is this all the luggage you have?’;
(ii) **Moshimoshi Akagi-san no otaku desuka**, meaning ‘Hello, is that (Mr.) Akagi speaking?’.

The ‘luggage’ sentence is a routine question usually heard at the customs office and the ‘hello’ sentence is a conventional expression at the beginning of telephone conversations. The sentences were selected because they are so commonly used and so conventional that the listener should pay minimal attention to the content. The scenarios used are:

1. A young customs officer talking to three types of passengers at the airport (a respectable gentleman, a casually dressed young student and a shabby drunk).
2. A young public officer talking to the same three types of citizens on the telephone.

There are two issues which require careful consideration in any investigation with or about the Japanese language. They are gender difference and regional accent difference. It is a well-known fact that Japanese male and female speech differs in the use of linguistic forms (Loveday, 1986) and also differs in the use of certain prosodic features such as pitch level in polite speech (Loveday, 1981). Therefore, it was considered wise to analyse the male and female speech separately. Male speech was focused on in the present study for two reasons. First, male utterances are generally lower in pitch, which makes acoustic analysis and synthesis easier than high-pitched utterances and second, Japanese men were reported to be more sensitive to politeness levels than women in some language use surveys (Minami, 1987, pp. 147–165). One reason why this gender difference arises may be that over polite utterances are associated with femininity (which is not desirable in many cases).

The second issue is the regional accent difference. The Japanese language has diverse regional accents which can be different from each other in various ways, including pitch accent, intonation and articulation. This study focuses on standard (or common) Japanese, which is based on the Tokyo dialect. Standard Japanese is taught at school and can now be heard in all parts of Japan together with regional dialects, due to the strong influence of television and the high rate of mobility of the population. Generally, people speak standard Japanese with their own regional accents, while some speak it without any trace of their accents and others retain some regional features, depending on their linguistic/educational backgrounds and willingness to adopt the standard (or Tokyo) accent. At the same time, it has become increasingly difficult to find perfect speakers of standard (or Tokyo) Japanese mainly because people frequently move in and out of the Tokyo area. According to a survey by the National Language Research Institute (Kokuritsu Kokugo Kenkyuujo) in 1981, more than half of the population in Tokyo came from other parts of Japan. This suggests that even people who were born and grew up in Tokyo are likely to have come in contact with various dialects. In the present study, 6 speakers from different parts of Japan were used. All the speakers could speak standard Japanese and there were no readily noticeable regional patterns in the recordings. The utterances were later assessed to be of standard (or Tokyo) accent by 3 Tokyo speakers (see Section 4.2). Future studies, however, need to consider highly localised regional samples for acoustic analysis of such fundamental factors as F0 contours. Acoustic analysis may be contaminated by regional differences.

A total of 6 untrained male native speakers of Japanese in their twenties participated in the recording sessions, each lasting about 2 hours. The speakers were given descriptions of the two situations and the characteristics of the speaker
(a customs officer in scenario 1 and a public officer in scenario 2) and the three types of addressees and scripts of short conversations between the speaker and the addressee. Although there is a strong association between the type of addressee and speaking styles in Japanese speech, the desired styles (i.e., ‘polite’ to a gentleman, ‘casual’ to a student and ‘authoritative’ to a drunk) were also specified in the instructions. The speakers, who knew each other well, worked in pairs, one playing the role of speaker and the other the role of addressee. After a practice session, the scripted conversations and the test passages (which the speakers were requested to speak two or three times) were recorded. They then exchanged roles and repeated the same procedures. The recordings were made on digital audio tape with a Sony TCD-D7 DAT recorder and then digitised at a sampling rate of 16 kHz onto a Sun workstation via its own A–D boards.

3. Acoustic analysis

3.1. Major acoustic variables in relation to affect

3.1.1. F0 variables

Mean or median F0 have commonly been used for indicating pitch level, whereas S.D. or coefficient of variation (S.D. divided by mean) and the difference between the peak F0 (highest or 90th percentile point) and the floor (the lowest or 10th percentile point) of the F0 contour have been used for pitch variability or range. F0 rate of change or slope has also been used to examine the speed of F0 movement (Fairbanks, 1940; Ross et al., 1986; Henton, 1995). In order to describe the shape of F0 contour and movement, several variables or indices have been used: the average number of changes in direction per second during phonation (Fairbanks, 1940; Ross et al., 1986; Henton, 1995). The recordings were made on digital audio tape with a Sony TCD-D7 DAT recorder and then digitised at a sampling rate of 16 kHz onto a Sun workstation via its own A–D boards.

3.1.2. Temporal variables

Speech rate (with or without pauses) has been widely used as an index of tempo and as a major acoustic correlate of tempo (van Bezooyen, 1984, p. 64). The other variables commonly measured are: total utterance duration, silence duration (Cosmides, 1983; Ogino and Hong, 1992) and the ratio of pause duration to speech duration (Ross et al., 1986). The importance of rhythm or microtemporal structure is also acknowledged by researchers (Miller et al., 1984; Brown and Bradshaw, 1985), but few studies have focused on this, perhaps reflecting the difficulty of defining rhythmicity.

3.1.3. Intensity variables

Several variables associated with intensity have been measured in relation to affect: amplitude level, variability, rate of change (Ross et al., 1986), amplitude fall–rise patterns (Cosmides, 1983) and peak values in phrases (Ogino and Hong, 1992).

3.1.4. Variables associated with voice quality and articulation

In addition to pitch, tempo and loudness, the effects of voice quality and articulation have also been investigated. Auditory assessment has been used to study voice quality in relation to introversion, dominance, sociability and emotional stability (Mallory and Miller, 1958; Scherer, 1979b). However, this approach has been criticised for the impressionistic labels used and a search for objectively measurable variables has been undertaken. Since Ogino and Hong’s (1992) survey showed that Japanese polite speech was considered to be associated with ‘soft’ or ‘lax’ voice, we focus on the ‘tense’/lax voice types here. The acoustic variables investigated are: the relative amount of energy in the upper harmonics (van Dusen, 1941; Frolova-Jensen and Prytz, 1976, p. 3), the shape of glottal waveform and the spectral energy distribution (Monsen and Engebretson, 1977; van Bezooyen, 1984, pp. 61–63; Childers and Lee, 1991). The lax voice often co-occurs with
breathiness. The potential acoustic variables for breathiness are the damping characteristics of the wave (Laver, 1980, p. 127; Childers and Lee, 1991), the amplitude difference between the lowest harmonic (F0) and the higher harmonics (Bickley, 1972; Henton and Bladon, 1985) and noise-to-harmonic ratio (Childers and Lee, 1991; de Krom, 1994; Klasmeyer and Sendlmeier, 1995).

Articulation (e.g., precise or careless) is intuitively important for expressing politeness because of a link with carefulness. Acoustic characteristics related to the articulation type are formant trajectories (Scherer, 1982, pp. 162–163) for vowel quality, voice onset time of stop consonants (Fairbanks, 1940) and intensity of release for stop consonants (Williams and Stevens, 1972) for consonant quality.

3.2. Measurement of acoustic features

F0 and temporal variables form the focus of this study for several reasons: first, they are major acoustic correlates of perceptual pitch and tempo; second, they are robust in the sense that they survive even in very noisy environments and through degraded telephone lines; third, they are relatively easy to measure and manipulate by means of currently available computer software; finally, they are thought to be important by Japanese people in politeness judgements in Japanese speech (Ogino and Hong, 1992).

Among F0 contour variables, the final F0 movement was particularly focused on in the present study, because there is greater freedom in F0 movement at the end of the sentence compared with other parts of the sentence in Japanese. In Japanese, pitch accent is fundamental to defining word identity; thus, speakers of Japanese have very limited freedom in terms of the pitch pattern (i.e., the position of the rises or falls of pitch) of an utterance, with the exception of the end of a sentence, which usually carries a final sentence intonation. In other words, sentence intonation in Japanese does not change the pitch pattern of the utterance except the final vowel, on which different types of tone (i.e., rising, level or falling) can be imposed. Furthermore, the final part is clearly important for expressing the speaker’s state, since particles signalling attitudinal meanings of speakers are generally located at the end of a sentence in Japanese. The importance of the final part as an information carrier is seen too in Sichuanese Mandarin in expressing affec (Chang, 1958), in distinguishing contrasted intonation in American English (Takefuta, 1972) and in controlling the flow of discourse in English (Brazil et al., 1980). So the final part has importance too in languages other than Japanese.

Selected utterances from the recordings (which are described in Section 1) were used for acoustic analysis. They were utterances of two sentences (i.e., luggage and hello) spoken by 6 male Japanese speakers in two speaking styles: casual and polite. Among utterances spoken by the same speaker for the same style and sentence, the first utterances were used because they sounded most natural; speakers tended to ‘recite’ their subsequent utterances. Utterances which were meant to be authoritative to a ‘drunk’ were excluded from this analysis, because most speakers had difficulties using the desired tone, tending to adopt either a seemingly casual or soothing tone. The question was whether distinct acoustic variables could be identified as consistently distinguishing these two different speaking styles.

F0 values and segmental durations were measured using the digital signal processing software package ESPS/Waves (Entropic Research Laboratory, 1993). The two sentences used in this acoustic analysis can be divided into two parts: the luggage sentence consists of ‘nimotsu-wa’ (phrase 1) and ‘koredake desuka’ (phrase 2); the hello sentence consists of ‘moshimoshi’ (phrase 1) and ‘Akagi-san no otaku desuka’ (phrase 2). A pause could be inserted between phrases 1 and 2. Acoustic variables analysed were mean F0 (in Hz), coefficient of variation, range (the difference between the 95th percentile point and the 5th percentile point in semitones), rate of change of F0 (mean for regression coefficients fitted for manually smoothed F0 contours of each vocoid segment), total length of the utterance (in ms), the duration of the final vowel (in ms), F0 direction (rise, fall or level) and speech rate excluding pauses and the final consonant and vowel
(in mora\(^1\) per second). Although there may be other potentially relevant variables for politeness levels, these variables were selected for the following reasons:

1. they have been investigated in relation to speaker variables such as personality attributions and emotions (see Section 3.1);  
2. some of them (e.g., speech rate and pitch level) were acknowledged to be relevant by native speakers in Ogino and Hong’s (1992) survey;  
3. they were relatively easy to manipulate by using digital resynthesis.

The duration of final mora in phrases were excluded for calculation of speech rate, because great variability was found among them. The regression coefficients were calculated by using normalised F0 values for each speaker. All F0 values of each speaker were normalised in such a way that the lowest F0 and the highest F0 of the speaker is 0 and 100. The lowest and highest F0 were among F0 values of all the utterances of the two sentences and a 1–2 minute dialogue with the speaker’s recording partner in the recording session.

Results of the acoustic analysis are shown in Tables 4–6. Table 1 shows the comparison between F0 variables in the polite and casual styles by 6 speakers. Great variability in use of these variables is apparent among the 6 speakers. For example, speakers TN and KS adopted higher pitch for the casual style, while HA used higher pitch for politeness and SF did not change pitch level very much. TK adopted a higher pitch level for the casual style in the luggage sentence, but lower pitch level in the hello sentence. This variability appears to suggest that the importance of these F0 variables as a cue to politeness is inconclusive. However, it does not necessarily mean that F0 variables cannot be one of signals for politeness; speech production varies from speaker to speaker depending on his/her ability of realising politeness levels. On the other hand, the temporal variables (i.e., speech rate, total utterance duration and final vowel duration) are consistent across the 6 speakers (Tables 1 and 2). For example, all speakers adopted slower speech rate and longer total utterance duration (with a few exceptions) for the polite style. Furthermore, the mean values for speech rate and total utterance duration across the 6 speakers were significantly different, at least in one sentence, in these two different speaking styles. Therefore, these temporal variables could be an important cue for politeness.

4. Experiment: the role of final F0 movement and speech rate

4.1. Introduction

This experiment was conducted to investigate the roles of final F0 movement and speech rate in signalling politeness. Two sets of stimuli were presented to each listener-judge in the same rating session; one for investigating the role of final F0 movement and the other for speech rate. An analysis using politeness scores of the former set of stimuli is referred to as experiment A and an analysis using the scores of the latter set of stimuli as experiment B.

4.2. Selection of acoustic features, manipulation techniques and source utterances

Two acoustic features, the final F0 movement (i.e., the duration and F0 direction of the final vowel) and speech rate (i.e., the overall tempo of the utterances), were selected on the basis of the acoustic analysis (Section 3.2). These two features were found to be consistently differently used by speakers in differentiating the two speaking styles (i.e., polite and casual), therefore, they appear to be relevant to the degree of politeness conveyed by these two styles.

It is true that intonation, especially the sentence’s final intonation which is mainly realised by the final F0 movement, is related to the meanings/functions of an utterance and in some cases it plays a key role in defining the main function of the utterance (e.g., ‘question’ versus ‘statement’).

---

1 Japanese has been described as a ‘mora-timed’ language by linguists. The unit ‘mora’ is defined as a particular sequence of segmental units: a consonant followed by a vowel (CV), a vowel, a ‘moraic’ nasal (/N/) or the ‘moraic’ first part of a long consonant (/Q/).
In spite of its close relationship with the meaning, the direction of a final tone was selected as a variable in the present experiment. It was because both final-rises and final-falls were actually used by native speakers in the same situation (see Table 1). In the case of the sentences used in the recordings (Section 2), the final F0 direction (i.e., rising or falling) did not greatly change the function and the core meaning of the utterances; the function of the sentences can be categorised as ‘confirmation-question’ and a moderate final-rise emphasises that the statement is a question whereas a moderate final-fall does not carry this emphasis.

---

**Table 1**
Comparisons between F0 and temporal aspects of polite versions (P) and those of casual versions (C) of two sentences spoken by 6 male speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>ST</th>
<th>F0 Mean (Hz)</th>
<th>S.D./mean</th>
<th>Range 95–5% (semitones)</th>
<th>Rate of changeb</th>
<th>Duration</th>
<th>Speech rate (mora/s)</th>
<th>Total length (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>C</td>
<td>P</td>
<td>C</td>
<td>P</td>
<td>C</td>
<td>P</td>
<td>C</td>
</tr>
<tr>
<td>TN</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>KS</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>HA</td>
<td>1</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>TK</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>SF</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>KI</td>
<td>1</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
</tbody>
</table>

*a ST 1: luggage sentence; ST 2: hello sentence; <, > and ? (the difference is less than 5% of the minimum value of the two) show the relationship between the value for the polite version (P) and the value for the casual version (C) spoken by the same speaker.

*b Rate of change was assessed by mean values for absolute values of regression coefficients.

---

**Table 2**
Comparisons between the characteristics of the final mora of the first and the second phrase of the sentence of polite versions (P) and those of casual versions (C) of two sentences spoken by 6 male speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>ST</th>
<th>Duration (mora/s)</th>
<th>F0 rate of change (semitones/s)</th>
<th>Mora: wa/shi (in phrase 1)</th>
<th>Final vowel: a (in phrase 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>C</td>
<td>P</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>TN</td>
<td>1</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>KS</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>HA</td>
<td>1</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>TK</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>SF</td>
<td>1</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&lt;?</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>KI</td>
<td>1</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
</tbody>
</table>

*a ST 1: luggage sentence; ST 2: hello sentence; <, > and ? (the difference is less than 5% of the minimum value of the two) show the relationship between the value for the polite version (P) and the value for the casual version (C) spoken by the same speaker.
The style difference was produced by human speakers’ subjective manipulation and F0 and segmental durations were manipulated by a computer program based on the TD-PSOLA technique (Charpentier and Stella, 1986; Moulines and Charpentier, 1990). Computer cue manipulation allows experimenters great control of major acoustic variables including F0 values and duration, but at the same time, this great flexibility could raise an ecological validity question: whether or not the manipulated speech remains natural or realistic? Manipulations could be too extreme to be natural or the interaction between the changed variables and the other variables might result in an unexpected impact on human perception (Scherer, 1979b, pp. 185–186; Brown and Bradshaw, 1985). In fact, it would be extremely rare that only one or two variables are changed while the others are kept constant in real speech, because human speakers produce sound by moving vocal organs, not by directly producing or modifying digital signals as computers do. Changing more than one variable, however, is not a problem if the process is a systematic one. The real problem is that the changes are not necessarily systematic or consistent across different speakers. Besides, in order to obtain a general idea on the effects of certain variables, it is vital to change the variables studied and keep the others intact. For these reasons, we decided to use computer cue manipulation techniques instead of human manipulation despite the ecological validity problem mentioned above.

Utterances of a single sentence (luggage) spoken by 2 male speakers (KS and TK) were used in this experiment. The luggage sentence was selected because utterances of this sentence had more variability in speech rate. In order to determine which of the 6 speakers, who took part in the recording sessions, to use, the utterances were assessed from a point of view of regional phonetic features (mainly pitch accent and intonation) by 3 female Tokyo speakers (who were born and raised in Tokyo) in their 20s, 40s and 60s. All of the utterances were judged as those of the standard (Tokyo) accent except one utterance. One judge found that the final intonation of one of the sentences (by Speaker HA in Tables 1 and 2) was ‘slightly different from the standard Tokyo usage’. Then another utterance evaluation test in terms of politeness degrees was carried out. Five subjects, who were native speakers of Japanese (3 females and 2 males in their 20s), rated all polite utterances and casual utterances separately, with a pairwise comparison method. The selected speakers, KS and TK, were judged as the best speakers, differentiating polite and casual versions clearly.

4.3. Experiment A: the role of final F0 movement

4.3.1. Design

A factorial $2 \times 2 \times 2 \times 2$ design was used with 2 speakers, two types of styles (polite and casual), two levels of duration of the final vowel (short and long) and two types of F0 direction (rise and fall).

4.3.2. Speech materials and stimulus preparation

The polite and casual versions of the luggage sentence spoken by 2 speakers (KS, from Kyushu, the southern island, and TK from Yokohama, in the eastern part of Japan) were used as source utterances. F0 and duration of the vowel in the final mora ‘ka’ in ‘Nimotsu … KA’, a sentence-final particle indicating that the sentence is a question, were measured in order to determine the values used for the following cue manipulation. Eight patterns were resynthesised based on the polite/casual version of the source utterance spoken by each speaker by manipulating F0 and duration of the final vowel ‘a’. The three variables were: two styles (polite and casual), two durations of the final vowel (short and long) and two types of F0 movement of the final vowel (rise and fall).

The final vowel of the polite version of speaker KS had an initial F0 point of 110 Hz and decreased by 54 semitones/s over 120 ms, while the TK final vowel had an initial F0 point of 110 Hz and increased by 28 semitones/s over 70 ms. KS’s final vowel of the casual version started from 117 Hz and slightly decreased 9 semitones/s over

---

2 A number of features, which include articulation, pitch, rhythm, loudness and voice quality, contribute to form speaking style differences. The term (speaking) ‘style’ is used as a label for referring to these features responsible for speaking style differences (i.e., polite and casual).
two-thirds of the total duration of 240 ms and then increased at the rate of 56 semitones/s over the rest; TK’s final vowel started from 101 Hz and increased by 20 semitones/s over 130 ms.

Two types of the final duration variable were set to the actual final duration of each speaker’s polite and casual versions (i.e., 120 ms for KS and 70 ms for TK for the ‘short/polite’ duration and 240 ms for KS and 130 ms for TK, for the ‘long/casual’ duration). It should be noted that the labels short and long mean that the durations labelled as ‘short’ are relatively shorter than the durations labelled as ‘long’. For the F0 contour, a straight-line was used for all the versions, except the long versions by KS, for which two straight-lines were used (i.e., a slightly falling one and a rising one) in order to simulate naturalness. The acoustic analysis of the source utterances showed that the rate of change of the final vowel was between 20 and 30 semitones/s for TK and between 50 and 60 semitones/s for KS, but the latter very steep fall/rise versions were judged somewhat less natural in a pilot study, when they were resynthesised. Therefore, ±25 semitones/s was adopted as the rate of change with the initial F0 points of the source utterances. The acoustic manipulation for KS’s polite version of the final vowel is shown in Fig. 1.

There were 6 occurrences for each stimulus condition. These stimuli were mixed with the stimuli for the experiment on the speech rate (experiment B) in random order. A total of 164 stimuli, consisting of 6 occurrences for a total of 26 conditions (16 conditions for experiment A and 10 conditions for experiment B) and a total of 8 dummy stimuli at the beginning and the end of three sub-sessions, were recorded on high-quality audio cassette tapes in random order. Each utterance was preceded by a warning tone and followed by a 3-s silence during which subjects were asked to make ratings.

4.3.3. Rating scale

A bipolar 8-cm scale of politeness was used, with ‘totemo teinei-denai’ (‘very impolite’) and ‘totemo teinei’ (‘very polite’) on the extremes. Japanese politeness (‘teinei-sa’) is closely associated with the relative relationship between speaker and addressee. The polite utterances whose addressees were respectable gentlemen were expected to receive plus scores, while the casual utterances whose addressees were young students were likely to receive minus scores on this scale.

4.3.4. Subjects

Twenty paid subjects (12 males and 8 females), mostly students from two universities in the eastern part of Japan, participated in the experiment. They were all native speakers of Japanese aged between 20 and 36. 7 male and 5 female subjects were from the eastern part of Japan, 4 males and 2 females from the western part and 1 male and 1 female from other parts of Japan.

4.3.5. Rating sessions

Subjects were given written instructions in Japanese, telling them that they would hear only one Japanese sentence spoken in various ways and their task was to rate utterances on a scale of politeness according to their own evaluation criteria. They were also informed that the speakers were young customs officers trying to speak politely to a respectable gentleman and also rather casually to a young student.
At the beginning of the session, the subjects were presented with polite and casual source utterances by these 2 speakers in order for them to assess the range of politeness of the stimuli. In a practice session they listened to 6 stimuli including two utterances with the maximum degrees of manipulation used in this experiment and rated them on the politeness scale. They then listened to the 164 stimuli in random order over Sennheiser HD 480II headphones in three sub-sessions with a short break between them. At the end of the session the subjects' speech rate was assessed; this procedure is described in experiment B. The 20 subjects were tested individually in a quiet room, each session lasting about 40 min.

4.3.6. Results and discussion

The politeness scores were obtained by measuring the distance between subjects' markings on the linear scale and a mid point on the scale. Scores could range between −4 (very impolite) and +4 (very polite). Kendall’s coefficient of concordance ($W$) was calculated to assess the general agreement among 20 listener-judges’ ratings of the four conditions (i.e., the combinations of short and long final vowel duration and rise and fall final vowel F0 directions) of both polite and casual styles originally spoken by the 2 speakers, KS and TK: $W = 0.53$ for KS’s polite style, $W = 0.57$ for TK’s polite style ($p < 0.0001$) and $W = 0.12$ for the casual style of both speakers ($p < 0.05$). The ratings for the polite style were more consistent than those of the casual style.

The mean values of 6 scores for each condition rated by each subject were calculated and shown in Table 3. An ANOVA was carried out on the mean values with factors of speaker (2 speakers), speaking style (polite or casual), final vowel duration (long or short) and final F0 direction (rise or fall). To assess the relative magnitude of the effects of each factor, we also calculated Eta-squared, a statistic based on the ratio of the sums of squares of a factor to the sums of squares of the total of all the within-subject factors. This statistic points to a possible salience difference; high scores in Eta-squared means that the magnitude of the effect of the factor is greater than low-scored factors. This analysis showed significant main effects of each factor apart from ‘style’, i.e., speaker ($F(1, 19) = 7.48$, $\text{Eta}^2 = 8.2\%$), final vowel duration ($F(1, 19) = 34.73$, $\text{Eta}^2 = 11.2\%$), final F0 direction ($F(1, 19) = 9.38$, $\text{Eta}^2 = 1.2\%$); interactions between

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Condition</th>
<th>Final duration</th>
<th>Final direction</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>Polite</td>
<td>Short</td>
<td>Rise</td>
<td>0.49</td>
<td>1.035</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Short</td>
<td>Fall</td>
<td>0.18</td>
<td>1.241</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Long</td>
<td>Rise</td>
<td>−0.97</td>
<td>1.544</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Long</td>
<td>Fall</td>
<td>−1.27</td>
<td>1.464</td>
</tr>
<tr>
<td>KS</td>
<td>Casual</td>
<td>Short</td>
<td>Rise</td>
<td>0.36</td>
<td>1.837</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Short</td>
<td>Fall</td>
<td>0.31</td>
<td>1.732</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Long</td>
<td>Rise</td>
<td>−0.63</td>
<td>1.874</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Long</td>
<td>Fall</td>
<td>−0.73</td>
<td>1.755</td>
</tr>
<tr>
<td>TK</td>
<td>Polite</td>
<td>Short</td>
<td>Rise</td>
<td>1.53</td>
<td>1.028</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Short</td>
<td>Fall</td>
<td>1.23</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Long</td>
<td>Rise</td>
<td>0.68</td>
<td>1.158</td>
</tr>
<tr>
<td></td>
<td>Polite</td>
<td>Long</td>
<td>Fall</td>
<td>0.06</td>
<td>0.988</td>
</tr>
<tr>
<td>TK</td>
<td>Casual</td>
<td>Short</td>
<td>Rise</td>
<td>0.45</td>
<td>1.143</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Short</td>
<td>Fall</td>
<td>−0.06</td>
<td>1.149</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Long</td>
<td>Rise</td>
<td>0.12</td>
<td>1.093</td>
</tr>
<tr>
<td></td>
<td>Casual</td>
<td>Long</td>
<td>Fall</td>
<td>−0.06</td>
<td>1.297</td>
</tr>
</tbody>
</table>

*a Style differences were produced naturally, by instructing the speakers to vary style.*
speaker and style ($F(1, 19) = 11.02, \text{Eta}^2 = 3.3\%$), final vowel duration and speaker ($F(1, 19) = 10.22, \text{Eta}^2 = 1.4\%$) and final vowel duration and style ($F(1, 19) = 25.54, \text{Eta}^2 = 1.4\%$) were also significant ($p < 0.05$). No interaction between duration and F0 direction of the final vowel was found, which suggests that these two acoustic variables function independently. According to Eta-squared, the factor of final vowel duration was the most salient factor with the F0 direction factor being less influential. The importance of the temporal aspects of the final part of sentences has also been acknowledged in (Imaizumi et al., 1994). They studied the effects of temporal variables in relation to emotions from the point of view of listener-adaptive characteristics in dialogue and found that the length of the final part of the target sentences ‘desuka’ played an important role in accounting for a factor which represented the emotional contrast between discomfort (e.g., awful, rough, etc.) and comfort (e.g., easy, kind, polite, etc.).

The mean values of the politeness scores for the short and long final duration versions for the polite and casual utterances (i.e., polite/casual versions of the source sentence) by the 2 speakers were calculated. The results are:

(a) The polite style: 0.34 for short and –1.12 for long for KS; 1.38 for short and 0.37 for long for TK.
(b) The casual style: 0.34 for short and –0.68 for long for KS; 0.20 for short and 0.03 for long for TK.

These differences, however, were not significant ($p = 0.05$). This shows that final rise was rated as more polite than final fall. However, the differences were less salient in comparison to those in the final duration of the utterance. Although the differences of mean values between the final rise and final fall versions are not very large, the majority of raters did prefer the final rise version to the final fall version in their politeness judgements: 52.5% of the subjects rated the final rise more positively than the final fall, while only 23.8% of the subjects preferred the final fall and 23.8% of the subjects showed no preference.

This final rise preference in relation to politeness may be related to an unmarkedness of sentence intonation, because the sentence used is a direct yes–no question whose unmarked intonation is a rising tone. This result agrees with the finding of Ogino and Hong (1992), who studied the final intonation of a sentence in Japanese in relation to politeness. They report that a level or falling tone was identified with polite versions for an expression ‘deshouka’, whose unmarked tone is a level or a slight fall (though no clear pattern was seen for non-polite utterances). Although politeness judgement is generally highly language-specific, the possible association between unmarked tones and politeness was also observed in (Scherer et al., 1984), who examined the final intonation of a German sentence in relation to several attitudinal meanings. They found that a final fall was rated more positively in wh-questions (i.e., who, what, when, where, why and how) sentences, while a final rise had higher ratings in yes–no question sentences on scales of agreeability and politeness. Since the unmarked intonation in wh-questions is a falling tone while that in yes–no questions is a rising one, they suggested that their results reflected the preference for the traditional description of ‘normal’ or ‘unmarked’ intonation.

Fig. 2 shows the mean politeness ratings for four conditions for each speaker (KS and TK). The four conditions are:

1. the speaker’s polite utterance with a ‘polite final prosody’ (i.e., the speaker’s short final duration and a rising F0 final direction);
2. the polite utterance with a ‘casual final prosody’
   (i.e., the speaker’s long final duration and a falling F0 final direction);
3. the casual utterance with the polite final prosody;
4. the casual utterance with the casual final prosody.

For both speakers’ utterances, the polite utterances with the polite final prosody were rated as more polite than their casual utterances with the casual final prosody (1-tailed t-tests, $p < 0.005$). KS’s casual utterance with the polite final prosody was rated as significantly more polite than his polite utterance with the casual final prosody ($p < 0.01$) and TK’s polite utterance with the casual final prosody did not sound more polite than his casual utterance with the polite final prosody. These results show that subjects were heavily influenced by the sentence-final prosody when they made politeness judgements. In fact, the final prosody had such a great effect that it almost overrode the speaker differences; when the casual final prosody was imposed on TK’s polite utterance (which, with the polite final prosody, was rated as clearly more polite than KS’s), this utterance became more or less the same as KS’s polite utterance with the polite final prosody.

Since Japanese regional accents are diverse, the effects of the rater’s own accent in relation to politeness judgements were also examined. The subjects were categorised into three accent groups:
1. eastern (Tokyo and its surroundings);
2. western (Osaka, Kobe, northern Shikoku and Hiroshima);
3. others (Toyama and Kagoshima).

Japanese mainland dialects were customarily divided into three language groups (Eastern Japan, Western Japan and Kyushu) (Shibatani, 1990, p. 196). Although it is acknowledged that the Japanese accent system distribution is highly complex, a special significance has been attached to the East–West division, because this division is seen not only in the linguistic domain but also in other socio-cultural domains (e.g., eating habits and the shapes of tools) (Tokugawa, 1981; Shibatani, 1990, pp. 196–202). Since politeness judgement is considered to be strongly influenced by listener-judges’ socio-cultural background as well as their linguistic background, this East–West division was adopted as a first step.

An ANOVA was performed with factors of speaker, style, final duration and final F0 direction as within-subjects factors and accent of the subjects as a between-subjects factor. The test showed a significant main effect of final duration and significant interactions between accent and style, speaker and style, and style and final duration ($p < 0.05$). The main effect of final F0 direction and the interaction between accent and speaker just failed to reach significance ($p = 0.06$). The mean ratings for the polite and casual utterances

Fig. 2. The mean values of politeness scores rated by 20 subjects for the polite/casual source utterances with the ‘polite final prosody’ (P-prosody) and ‘casual final prosody’ (C-prosody) in experiment A. The scores for the utterances by KS (a) and those for the utterances by TK (b) are shown separately.
spoken by TK, an eastern speaker and KS, a southern speaker, rated by the subjects from both the eastern and western part of Japan are shown in Fig. 3. It is interesting to observe that the eastern raters appear to have perceived the eastern speaker’s intention correctly, by rating TK’s polite utterances polite and his casual versions less polite, while the western raters failed to do so. As was mentioned above, since it may be too simplistic to use the East–West division as the accent category, caution is needed to generalise this result. However, the result does suggest the importance of controlling the accent factor of both speakers of stimulus utterances and listener-judges in politeness research.

Although the importance of the accent factor is undeniable, the results of the ANOVA test mentioned above showed that there was no significant interaction between the accent factor and the factors of any acoustic variables studied here (i.e., the final duration and final F0 direction) at the level of 0.05. In other words, subjects showed a difference in style preference, whereas they responded to the differences in the acoustic variables in the same manner. This may suggest that though the listeners’ own accent is important in making politeness judgements, the accent does not interact with the effects of these acoustic variables in terms of their ratings. This point is important, because, since these acoustic variables are manipulatable both by a speaker and by computer, this means that we can control politeness levels to some extent relatively easily and that this can be done independently from accent considerations.

4.4. Experiment B: the role of speech rate

4.4.1. Introduction

This experiment was designed to investigate the effect of speech rate on politeness judgements. Speech rate has been studied in relation to various kinds of attitudinal meanings including politeness. Brown et al. (1974) used 15 paired opposite adjectives, which were later clustered into three factors based on the patterns of the scores, as rating scales: benevolence including polite, ‘kind’, ‘sincere’, ‘religious’ and ‘just’, competence including ‘active’, ‘ambitious’, ‘intelligent’ and ‘confident’, with other scales including ‘happy’, ‘good-looking’, ‘strong’, ‘dependable’, ‘sociable’ and ‘likeable’. A series of studies found that first, increasing rate increased competence and decreased benevolence (Brown et al., 1974; Smith et al. 1975); second, benevolence had an inverted U-shape in function of rate, i.e., having the maximum scores at the normal rate (Brown et al., 1974; Smith et al., 1975; Brown, 1980), with American subjects; while Brown et al. (1985) found that decreasing rate linearly increased benevolence with British subjects; finally, rate manipulations may have had much greater and consistent effects than F0 mean and variation manipulations, especially for competence factors, but less clear effects for benevolence (Brown et al., 1974; Smith et al., 1975).

Most of the studies mentioned above used computer rate manipulation, which linearly
compressed or expanded each segment of the utterances and some researchers expressed their concerns about undesirable effects caused by this linear change in segmental duration (Apple et al., 1979; Brown, 1980), because rate change, which actually takes place in human speech, is a complex process. Articulation changes such as elision and assimilation, pause insertion and changes in spectral characteristics also take place along with rate changes (Campbell, 1992, p. 198). Further, consonants are more resistant to durational change than vowels and the position of a phoneme in an utterance cannot be overlooked when durational changes are made. Bell-Berti et al. (1995) report that when subjects attempt to speak slowly, this differentially affects early and late vowels, with the early vowels becoming stretched and a small effect on the late ones. Unfortunately, no satisfactory rules for predicting segmental durational changes as a function of speech rate in Japanese exist at the present time. Therefore, we adopted a linear compression algorithm in order to implement durational changes. An encouraging finding for using computer manipulation comes from Brown's (1980) experiment on the comparison between the effects of human alterations in rate and the effects of computer manipulations of rate. He asked 6 subjects to rate human manipulated versions and computer manipulated versions on the same rating scales used in (Brown et al., 1974) and found high correspondence between the human- and computer-manipulation results, though with uniformly higher benevolence ratings for the human manipulated stimuli.

4.4.2. Design

A factorial $2 \times 5$ design was used with 2 speakers (KS and TK) and five different speech rates (change rate of segmental duration of the source utterances: 0.8, 0.9, 1.0, 1.1 and 1.2).

4.4.3. Speech materials and stimulus preparation

The utterances were politely spoken by the slower speaker KS (speech rate: 10.8 mora/s) and by the faster speaker TK (speech rate: 11.8 mora/s). Rate was manipulated by means of computer resynthesis based on the PSOLA technique. All the unmodified versions were resynthesised with a rate change factor of 1.0 in order to preserve the factor of degradation caused by the PSOLA manipulation at the same level as that for the other rate versions.

4.4.4. Rating sessions

The politeness ratings were those already reported in experiment A, that is with 164 stimuli (6 occurrences each of the 26 conditions plus 8 dummies). The speech rate of the subjects was assessed at the end of the session; speakers were presented with written utterances in Japanese meaning "Hello, how do you do. I'm afraid I haven't brought anything to write with, could I borrow the ball-point pen over there?" and asked to speak to the experimenter as naturally as possible, in front of a small microphone on the desk.

4.4.5. Results and discussion

The politeness scores were obtained by measuring the distance between subjects' markings from a mid point on the bipolar scale. Scores could range between plus and minus 4. The mean values of 6 scores for each condition rated by each subject were calculated and an ANOVA test was performed. Eta-squared was also calculated to assess the relative weight of contribution of each factor. This analysis showed significant main effects of speaker ($F(1,19) = 23.73$, $\text{Eta}^2 = 28.5\%$) and rate ($F(4,76) = 11.99$, $\text{Eta}^2 = 15.7\%$) and a significant interaction between these two factors ($F(4,76) =$
Eta-squared showed that the main effects were stronger than the interaction, while the speaker factor was more salient than the rate factor. The mean values of politeness scores across 20 subjects for the five different versions of the sentence are shown in Fig. 4 as an inverted U-shaped function of rate.

Kendall's coefficient of concordance ($W$) was calculated to assess the general agreement among 20 listener-judges' ratings of the five different rate versions: $W = 0.47$ for KS's utterances and 0.40 for TK's utterances ($p < 0.0001$). Although we obtained a significant main effect of rate, the consistency among subjects' judgements was not very high. Therefore, we examined the speech rate of subjects in relation to individual rate preference, and obtained an interesting relationship between the speech rate of the subjects and their rate preference, as detailed below.

We calculated the speech rate of the subjects by measuring the speaking time of the utterance recorded at the end of the rating sessions. The speech rates of the 20 subjects showed a significant sex difference: the female subjects ($N = 8$) spoke significantly more slowly than the male subjects (2-tailed $t$-test, $p < 0.01$). It is interesting to speculate that this sex difference in rate might reflect a social expectation that women should be more polite than men, as well as a social expectation that slower speech is associated with politeness as Ogino and Hong's (1992) survey showed. We categorized the speech rate of the subjects into three groups for male and female subjects separately using a cluster analysis: slow-speaker (2 males and 4 females), middle-speaker (4 males and 3 females) and fast-speaker (6 males and 1 female). An ANOVA with factors of speech rate of listener and sex of listener as the between-subject factors, and rate of utterance and speaker of utterance as the within-subject factors, showed significant main effects of speaker and rate, and significant interactions between speech rate of listener and rate of utterance and a significant three-way interaction between speech rate of listener, sex of listener, and rate of utterance ($p < 0.001$). The important effects are summarised in Fig. 5, which illustrates an interesting relationship between the speech rate of the 12 male subjects and their rate preference. Slow-speakers rated slower utterances as more polite than middle/fast-speakers, while the middle/fast-speakers preferred faster versions. The female subjects' data, however, showed no clear difference between the slower-speaker group and the faster-speaker group and are not plotted. One possible account for this gender difference is that the female subjects might have used their 'formal voice' and, as a result, the speech rates might not reflect their own speech rate in everyday life.

5. Conclusion

The major findings may be summarised as follows. Acoustic analysis of the samples of polite and casual utterances showed that F0 movement of the final vowel of the sentences and their overall tempo were used consistently by all 6 speakers to indicate polite or casual situations, while F0 level, F0 range and F0 rate of change were not used consistently by the speakers to indicate levels of politeness. In order to confirm the effects of two variables (i.e., final F0 movement and speech rate) which were used differently according to the speaking styles, a perceptual experiment with
listeners was conducted. The results showed that the prosody (especially the duration) of the final vowel of the sentence had a great impact on politeness judgements; prosody information through the last 100 ms or so changed the total impression of the utterance. Speech rate was also found to be relevant. The main factors of speech rate and speaker, and the interaction between them were significant, and the function relating politeness and speech rate was that of an inverted U-shape.

Table 4
Comparisons between F0 variables of polite versions (P) and those of casual versions (C) of two sentences spoken by 6 male speakers

<table>
<thead>
<tr>
<th>SP</th>
<th>ST</th>
<th>P</th>
<th>C</th>
<th>S.D./mean</th>
<th>P</th>
<th>C</th>
<th>F0 range 95–5% (semitones)</th>
<th>Mean for regression coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN</td>
<td>1</td>
<td>147</td>
<td>155</td>
<td>0.29</td>
<td>0.35</td>
<td>16</td>
<td>19</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>163</td>
<td>166</td>
<td>0.27</td>
<td>0.30</td>
<td>15</td>
<td>20</td>
<td>2.16</td>
</tr>
<tr>
<td>KS</td>
<td>1</td>
<td>147</td>
<td>162</td>
<td>0.24</td>
<td>0.25</td>
<td>16</td>
<td>14</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>171</td>
<td>178</td>
<td>0.21</td>
<td>0.25</td>
<td>12</td>
<td>13</td>
<td>1.03</td>
</tr>
<tr>
<td>HA</td>
<td>1</td>
<td>146</td>
<td>136</td>
<td>0.17</td>
<td>0.11</td>
<td>9</td>
<td>5</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>156</td>
<td>139</td>
<td>0.15</td>
<td>0.20</td>
<td>10</td>
<td>11</td>
<td>1.14</td>
</tr>
<tr>
<td>TK</td>
<td>1</td>
<td>124</td>
<td>143</td>
<td>0.24</td>
<td>0.18</td>
<td>12</td>
<td>10</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>158</td>
<td>149</td>
<td>0.13</td>
<td>0.13</td>
<td>8</td>
<td>7</td>
<td>2.4</td>
</tr>
<tr>
<td>SF</td>
<td>1</td>
<td>134</td>
<td>134</td>
<td>0.28</td>
<td>0.27</td>
<td>15</td>
<td>13</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>145</td>
<td>148</td>
<td>0.19</td>
<td>0.15</td>
<td>11</td>
<td>10</td>
<td>2.5</td>
</tr>
<tr>
<td>KI</td>
<td>1</td>
<td>115</td>
<td>113</td>
<td>0.21</td>
<td>0.28</td>
<td>10</td>
<td>11</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>122</td>
<td>129</td>
<td>0.29</td>
<td>0.22</td>
<td>9</td>
<td>10</td>
<td>2.9</td>
</tr>
</tbody>
</table>

*ST 1: luggage sentence; ST 2: hello sentence.

Table 5
Comparisons between temporal variables of polite versions (P) and those of casual versions (C) of two sentences spoken by 6 male speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>ST</th>
<th>P</th>
<th>C</th>
<th>Total utterance duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN</td>
<td>1</td>
<td>10.4</td>
<td>11.6</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.5</td>
<td>13.2</td>
<td>2.16</td>
</tr>
<tr>
<td>KS</td>
<td>1</td>
<td>10.8</td>
<td>11.9</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.9</td>
<td>13.2</td>
<td>2.00</td>
</tr>
<tr>
<td>HA</td>
<td>1</td>
<td>11.6</td>
<td>14.3</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13.5</td>
<td>13.9</td>
<td>1.88</td>
</tr>
<tr>
<td>TK</td>
<td>1</td>
<td>11.8</td>
<td>14.1</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12.7</td>
<td>13.0</td>
<td>1.53</td>
</tr>
<tr>
<td>SF</td>
<td>1</td>
<td>11.8</td>
<td>12.3</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13.5</td>
<td>14.1</td>
<td>1.50</td>
</tr>
<tr>
<td>KI</td>
<td>1</td>
<td>11.8</td>
<td>13.2</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12.8</td>
<td>13.0</td>
<td>1.49</td>
</tr>
</tbody>
</table>

*ST 1: luggage sentence; ST 2: hello sentence.
Since the consistency between listener-judges was not very high, the factor of the speech rate of listeners was examined. The positive correlation between the speech rate of listeners and their preferred rate of utterance was obtained. Although the rate factor was found to be significant, it was not powerful enough to override the speaker/style difference. Since there are usually clear differences between polite utterances and non-polite utterances, which native speakers could easily report, there must be other factors which are responsible for the speaker/stylistic differences, which we have as of yet been unable to identify. Further research in this area is clearly needed.

It is increasingly clear to us that listener characteristics must be carefully considered in politeness research. Although there was a high level of inter-judge agreement on the scale of politeness in the present study, we did find that characteristics such as the speech rate and regional accent of the subjects themselves had significant effects on their judgements. People appear to be very sensitive to unnaturalness by their standards and this listener-specific sensitivity may bias politeness judgements. A single extreme value for any acoustic feature (e.g., very fast speech rate) may reduce perceived politeness, but this will differ listener by listener.

Therefore, we conclude that polite utterances require that every influencing feature be kept within a certain range, which will vary from politeness level to politeness level, from speaker to speaker and indeed from listener to listener.

Acknowledgements

The authors thank Akiko Kawasaki of the University of Tsukuba and Machiko Sato of Tokyo Institute of Polytechnics for helping with the perceptual experiments. We are also grateful for detailed and most helpful comments from the anonymous reviewers.

References


