The contribution of voice quality to the expression of politeness: an experimental study

Mika Ito

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

Department of Theoretical and Applied Linguistics,

University of Edinburgh

2005

Abstract

This thesis investigates the role of voice quality in the expression of politeness under conditions of varying relative social status among Japanese male speakers. The thesis also sheds light on four important methodological issues: 1) experimental control of sociolinguistic aspects, 2) eliciting semi-natural spontaneous speech which satisfies naturalness, 3) recording quality suitable for voice quality analysis, and 4) the use of direct waveform and spectrum measurement as a non-invasive method for measuring glottal characteristics related to perceived voice quality.

Japanese has been believed to rely on what has been called "negative politeness" (*formality* and *deference*). Since explicitly expressing deference under the *Keigo* (Japanese system of honorifics) requires mastery of a highly complex system, in daily conversation, this function may be taken over by vocal paralinguistics. Also, as the *Keigo* system is not supposed to convey "positive politeness" (*friendliness* and *solidarity*), vocal paralinguistics may contribute to this aspect of politeness. The use of high fundamental frequency (F0), which has been considered a universal cue of politeness, is more associated with femininity in Japanese, and this usage of F0 is not observed in male speakers, who possibly employ other vocal cues to express politeness. Therefore this study focuses on voice quality of male speakers in expressing politeness.

To obtain natural, unscripted utterances, the speech data were collected with the Map Task. This task also allows us to study the effect of manipulating relative social status differences among participants in the same community. For voice quality analysis, direct waveform and spectrum measurement (Hanson 1995) was employed. We mainly

computed relative amplitudes of harmonics and formant peaks in the spectrum, as alternatives to certain well known parameters used in previous studies. We also measured F0 and amplitude perturbations as possible indicators of voice quality.

An experiment was conducted to observe the alignment between acoustic measures and the perceived politeness from both written and spoken versions of the utterances obtained from the Map Task. The results suggest two principal findings. First, *Keigo* does not play a role in conveying politeness in everyday conversation, but speakers showed politeness through voice quality variations. Second, in judging the politeness of test utterances, listeners reacted to the irregularity of the waveform and spectral characteristics in the third formant region. In particular, the speakers' spectral tilt range and correlation coefficient between ratings of politeness and spectral tilt are anti-proportional, and extremely large or small spectral tilt values were consistently perceived as not appropriate when addressing social superiors. These results are expected to contribute to both sociolinguistics and speech technology, such as speech synthesis.

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Mika Ito

Acknowledgements

In my PhD years, the University of Edinburgh has helped improve my knowledge towards linguistics. Since I have come from an engineering background, my linguistic background has been gradually strengthened by the excellent staff at the university. The department has a strong emphasis on linguistic theory and at the same time, the department accommodates technical research branches. I believe that this department is the best environment for carrying out my research. I would like to thank Anne Cutler, who introduced me to my principal supervisor and pointed me in the direction of linguistics and this department. Otherwise, a linguistic approach would never have occurred to me. Among all the university people, my supervisors have enlightened me to a better understanding of the various theories of speech analysis and linguistics. D. Robert Ladd, my principal supervisor, paved my linguistic approach towards this topic. He is a genuine linguist, who has deep insights and wide perspectives towards any phenomena he observes, and I have been stimulated by his viewpoint of the humanities. Alice Turk has sharpened my skills, especially speech analysis. And at the same time, she is very warm person at heart and she gave me a lot of encouragement at some crucial points.

The combination of linguistics and technology has led me to make great progress in speech analysis, considering sociolinguistic theory. I owe Miriam Meyerhoff and Mits Ota big thanks for leading me to recent articles to update my knowledge of Japanese politeness, especially from the theoretical point of view. The same gratitude is going to Haruo Kubozono.

In addition, I belong to the "P-workshop" organised by the department which enhances the communication between phonetics, phonology and speech-technology people, and I

had many suggestions from the members of P-workshop. I also note that the organisers of P-workshop, Cassie Mayo and Robert Clark offer formatting resources to us which are very helpful for presentations.

The departmental staff offers a powerful resource and network when it comes to offering an optimised environment for researchers doing speech technology. I owe a lot to Cedric MacMartin, Michael Bennett, and Eddie Dubourg for their support through my study.

As for psycholinguistic issues, the members of HCRC, above all Cathy Sotillo, informed me the essential methodology of the Map Task, since it was developed at HCRC.

I also owe a lot to Syun Tutiya, Yasuo Horiuchi, and Akira Ichikawa, who kindly offered the materials such as instructions, maps and detailed documentation of procedure of Japanese Map Task Corpus, by Chiba University. They encouraged me by showing how this kind of multidisciplinary project should be conducted in a cooperative manner.

For conducting the production and perception test in Tokyo, I would like to thank my undergraduate supervisor, Keikichi Hirose, who kindly allowed me access to his laboratory's facilities and staff (including him!) through these experiments which had been conducted for three spring seasons. I cannot imagine if I could carry out my project without their support.

I would like to thank Helen Hanson, John Laver, Christer Gobl, Ailbhe Ní Chasaide, Norma Antoñanzas-Barroso, Bruce Gerratt and Jody Kreiman who are specialists in voice quality and have been particularly helpful. They gave technical advice and comments on my work when I needed, and also kept me informed of their own recent work.

However, any mistakes that remain are my own.

I also would like to thank my lab mate, Irena Yanushevskaya, who kindly read my earlier versions of this work and her careful advice and comments through the period of writing up this thesis.

I would like to thank Aoju Chen, Yoko Matsumoto-Sturt, Sherry Ou, Mariko Sugahara, and Ivan Yuen for being nice company for me during conference journeys and study in Edinburgh. Also I would like to thank Akemi Iida, Mihoko Teshigawara, and Ken-Ichi

Sakakibara, for being nice company, kept on encouraging me during my PhD work, although they live afar. The same goes to Makoto Nomura, especially for his moral support during the revision process.

Next to last, I would like to thank my family, above all, my mother Mitsuko Ito and my grandmother Kazuko Funahashi, for their deep understanding and strong support through this study. And, if I inherited any good sense for studying sound and speech, it would be from the influence of my deceased father, Shuuki Ito, therefore my gratitude goes to him.

This PhD coursework was supported by Overseas Research Student Award by British Government and a Faculty of Arts Studentship from the University of Edinburgh.

Contents

Chapter 1

Introduction

1.1 Vocal paralinguistics and its role in the expression and perception of politeness

The Akutagawa award-winning novelist, Machida (1999) does not like the announcement of the trolley service on the *Shinkansen* (the bullet train). He writes that the expression "*karui o-nomimono, karui o-shokuji*" (= 'light snack and light drink') lingers in his mind. He thinks that the expression is odd because it contains "*o-*" (honorific prefix) for snack and drink even though they are *light* and the voice signals the light, informal impression. For him, it does not sound as if they take customer care seriously. He concludes that the mismatch between the honorific expression and light impression of the announcement gives a strange impression, and that this impression is confirmed by the sour taste of coffee served on the train.

I do not argue here about whether the coffee served on the bullet train is sour or not. There are two important issues that we discuss in this thesis. First, Machida pointed out that the honorific prefix used in the announcement was somewhat awkward, and did not work successfully to convey the politeness that the honorifics are supposed to convey. Second, he referred to the impression that the voice of the announcement corresponds

well with the situation, while the actual words of the announcement do not.

In daily life, we often rely on things that we can perceive in an utterance beyond what the speaker literally says. We try to get any information available not only from "on record" signals but "off record" signals (see discussion about politeness theory by Brown and Levinson (1987), which appears in Chapter 2). We even consider that quite often the truth lies in a nonverbal expression rather than a verbal expression.

Why does this happen and how? To investigate this, we need to examine closely what is going on between verbal and nonverbal expressions in real conversation. As the example at the beginning of this chapter shows, studying politeness is likely to give rich information regarding the parallels between linguistics and paralinguistics.

From situation to situation, the way of expressing politeness differs. For example, in Akihabara (the street in Tokyo which specialises in electronics stores), many sales persons are likely to use high pitch, and for more than one reason. They encourage people to purchase their products, they compete with background noise, and they express politeness by speaking at the top of their voice which is an unlikely way of expressing politeness in normal daily conversation. On the other hand, receptionists at customer support centres of electronics companies where they process complaints from customers are usually trained to use low pitch and "calm voice" (so called by the human resource and education section of one of such companies), even though they also belong to the section of "customer sales" of electronic products. In this way, the receptionists can give the impression that they are trustworthy and they can assure reliable support.[1] Even sales people use different voices according to the situation of a particular sale, because people's expectations differ.

[1]This information is based on my experiences while working as an employee for a Japanese electronics company.

Therefore, the study of politeness needs to be carefully designed taking into account the social situation, with a strong consideration of theory.

1.2 The aim of this study

In this thesis, the main focus is vocal paralinguistics, especially voice quality, in relation to the expression of politeness in our daily life. In particular, this study's aim is to determine whether and how Japanese speakers use breathiness to express/perceive politeness. For this purpose, I will discuss issues such as:

- Comparison of universal politeness strategy and Japanese politeness strategy as it is used in contemporary daily conversation.
- The two main strategies which are believed to be Japanese politeness: deference and formality, which are highly accepted and preferred in Japanese society, and therefore reflected in the Japanese honorific system, *Keigo*.
- The function of vocal paralinguistics compared to that of *Keigo* (the Japanese system of honorifics) in expressing politeness.
- Why voice quality, especially breathiness, should be studied in the context of politeness.
- The methodologies for studying breathiness quantitatively, in a minimally invasive manner.

On the other hand, I will NOT discuss the following issues in this thesis:

- The reason why the Japanese society has come to prefer the politeness strategy of formality and deference. Ordinary people may say, "It came from *Bushido* (the way of *Samurai*)," or, "We learned in Japanese history class that the Edo government introduced Confucianism together with rigid social class system, to discourage people's attempt to plan revolution." These arguments need to be examined rigorously by social historians. The focus of this study is what is going on currently in Japanese society, especially vocal paralinguistic ways of signalling politeness. I briefly discuss

the current situation of contemporary conversational usage of Japanese *Keigo* as an issue to be considered in the experimental design of this study, but not beyond this.

- The stereotype that "the Japanese are very polite people". This stereotype is widely believed among Japanese people, including traditional Japanese linguists, but this would be better explained in the context of preference of politeness strategy. For this purpose, I introduce the politeness theory proposed by Brown and Levinson (1987).
- Acting voice, which is likely to lead us only to the stereotype that speakers have in their mind, rather than how they would behave in a real situation. This study devoted considerable effort to avoid such effects, and to elicit semi-natural spontaneous speech with politeness.

To achieve the aims above, this study focuses on a particular situation which is likely to happen in our daily life. Focusing on one particular situation allows us to have experimental control over social factors by limiting the number of factors which influence each part of the experiment. This control should be thoroughly considered in experimental design. At the same time, effort should be made to ensure that the participants are not acting. This control is especially necessary in order to avoid the interference from stereotypes that recording participants have. Acting configurations should be avoided in experimental design, such as preparing scripts and asking participants to pretend that the addressee is there for him/her. This study aims at avoiding this acting methodology, which was employed by most past studies.

To avoid these problems, the experimental design should not rely on stereotypes but needs a strong basis in politeness theory. It is necessary to have a careful review of the theoretical perspectives on universal politeness and specific strategies that speakers use to achieve their aims in speech.

1.3 Structure of this thesis

The thesis consists of the following chapters.

Chapter 2 discusses theoretical views of politeness, and illustrates the strategy and expressions of politeness, especially linguistic and paralinguistic information, which is likely to be implemented in speech. We state that there are two issues which may cause confusion when discussing the term "politeness".

The first issue is that Westerners have in the past questioned whether the concept of politeness, as defined by Japanese people, is appropriately called "politeness". To clarify this problem, I will introduce the politeness theory proposed by Brown and Levinson (1987). According to this theory, there are many strategies available to express politeness, and Japanese politeness is based on some of these strategies.

The other source of confusion is *Keigo* (Japanese honorifics) and the understanding of *Keigo* put forth by traditional researchers of the Japanese language. To clarify the confusion between politeness and the function of *Keigo*, the following fact should be considered: the so called "polite-form (*teineigo*)" represents formality but not the whole notion of politeness as covered by Brown and Levinson (1987). Regarding this, the argument set by the traditional researchers of the Japanese language studies and the actual use of *Keigo* in Japanese society will be examined. If we examine closely how this theoretical and practical politeness is applied, we will be able to see the clear difference between the function of *Keigo*, the daily use of *Keigo* and vocal paralinguistics in expressing politeness.

After moving the focus to vocal paralinguistics, we discuss the difference in usage of suprasegmentals between genders. Through observations reported in recent literature, we will see that using high pitch (F0), which was regarded as one of the universal cues of politeness, is not an appropriate cue in the case of Japanese male speakers, normally

in their daily conversation, because high pitch serves as a cue to femininity and it is used to show politeness by female speakers of Japanese. We will also look for evidence that politeness can be expressed in voice quality, especially breathy voice, as a possible politeness indicator for Japanese male speakers.

In Chapter 3, the methodology of voice quality analysis will be discussed, as we investigate the role of voice quality, especially breathiness, as an indicator of politeness.

To distinguish different types of voice quality, while understanding the similarities and differences of voices, the taxonomy proposed by Laver (1980), which is introduced at the end of Chapter 2, will be revisited and examined more closely. Following the hypothesis that breathiness can be a cue for politeness, several characteristics of voice quality, and especially of breathiness, will be reviewed.

After giving a broad overview of voice characteristics, the findings of several measurements of glottal configurations related to perceived breathiness will be introduced. Owing to these findings, we can discuss the relationships among the following correlates: the physiological movement of vocal folds, aerodynamic and acoustic measures of acoustics, and perception of breathy voice.

As an important issue, the appropriateness of several measurement methods will then be discussed, for example, instrumental methods, source-filter decomposition with inverse-filtering and a direct waveform and spectrum measurement. As mentioned in section 1.2, observation of natural speech is one of the priorities in this study. After discussion of the advantages and disadvantages of each method, direct waveform and spectrum measurement (Hanson 1995) will be selected to be a good candidate to be implemented to meet this study's methodological objective. This is because it is a non-invasive method of estimating glottal configuration by measuring acoustic parameters. This makes it possible to assess with acoustic correlates of perceived breathiness while at the same

time preserve naturalness during the collection of speech.

Chapter 4 is the main body of this study. It consists of the description of two experiments; a production experiment and a perception experiment. Building on the background examined in Chapter 2 and 3, the experimental design of testing politeness and estimated voice quality will be introduced. This experimental design includes various strategies to improve experimental methods of testing the degree of politeness, expressed and perceived, which most of previous studies have not achieved.

In the production experiment, the following techniques are employed. For data collection, to elicit semi-natural spontaneous speech with high quality of recordings, the Map Task (Anderson et al. 1991) is employed. Among other things, this method enables us to set the configuration of speakers with different social status, which will help to reveal the intended degree of politeness more effectively.

In speech analysis, this study employs direct waveform and spectrum measurement as the main method of estimating perceived breathiness. As just noted above, this method has the great advantage of non-invasiveness. To allow for the fact that this method has some disadvantages - such as difficulty in estimating aspiration noise in the high frequency region from spectra of male speakers - second glottal excitation is also measured as a factor of consideration.

To aim at quantitative measurement of aspiration noise, F0 and amplitude perturbation quotients in the high frequency region are also introduced as alternative quantitative measures to noise rating method which is a part of the direct waveform and spectrum measurement.

In the perception test, to collect responses from subjects, the following techniques are employed together with a forced-choice test. To evaluate the differences of subtle nuances that *Keigo* may convey, Magnitude Estimation (Stevens 1969, Bard et al. 1996) is employed.

The results illustrate the complexities of human behaviour in terms of voice quality. The results from the production experiment and the perception experiment do not correspond with each other entirely. However, we will see some trends as follows.

First, the speakers change their voice quality in various ways so as to convey politeness.

Second, the listeners use spectral information in the high frequency region such as spectral tilt with consistency when they perceive politeness.

These findings lead us to conclude that vocal paralinguistics does play a role in conveying politeness.

Chapter 5 will summarise the results of the experiments, and will discuss the direction of vocal politeness studies to be conducted in the future.

This study's contributions are as follows. First, it demonstrates how the politeness theory proposed by Brown and Levinson (1987) can possibly be applied to the study of the Japanese politeness in speech, with the effects of *Keigo* taken into consideration. Second, owing to the waveform and the spectrum measurement employed in this work, the voice quality of natural spontaneous speech was investigated in a non-invasive way. Finally, the characteristics of "polite voice" revealed in this study also give some insights to speech applications. Using acting voice alone puts the emphasis on production side and this does not allow us to observe the strategies employed by speakers and listeners which, as this study revealed, can be quite different. This finding shows that we should develop speech applications from the listeners' viewpoint. The speech applications, such as speech synthesis and speech recognition, will benefit from this study to enable comfortable man-machine user interface.

Chapter 2

Politeness: Sociolinguistic perspectives

2.1 Introduction

Many researchers on Japanese state that *Keigo* - the lexical and grammatical system of honorifics - is the dominant way of expressing respect and deference in the language. It is also often claimed that respect and deference are the primary features of Japanese politeness, and that such features as friendliness and solidarity, which seem to form an important element of politeness in some Western cultures, play little role in Japanese. On the other hand, some studies of Japanese politeness demonstrate that Japanese listeners often use vocal paralinguistic cues for perceiving politeness even though lexical cues provided by *Keigo* may seem unambiguous, and that paralinguistic cues may even override lexical and grammatical cues. Some studies also seem to indicate that paralinguistic cues likely to convey friendliness can possibly be treated as a component of Japanese politeness. Thus there is a contradiction between the traditional assumption that *Keigo* is dominant and the results of experiments that study politeness based on vocal paralinguistic cues.

In this chapter I review some theoretical and empirical approaches to politeness, both in Japanese and cross-linguistically. The underlying goal is to examine the components of politeness - deference, respect, friendliness, etc. - and to review empirical evidence about

how these different components can be conveyed by different kinds of cues, such as lexical and vocal cues.

I will begin by describing the politeness theory of Brown and Levinson (B&L politeness theory), who view politeness from a universalist perspective. I will then discuss some Japanese scholars' counter arguments against the B&L politeness theory that come from problems specific to Japanese. One key problem is that politeness strategies are reflected in linguistic information in Japanese, especially in *Keigo*, the honorific system. This system is summarised briefly. This discussion will help us to understand the background of politeness strategies and expected attitude in the context of Japanese society. The complication of handling *Keigo* will demonstrate that taking into account relative social status and social/psychological distance is important in expressing politeness. Also, by observing the difference in politeness strategies between Western and Japanese culture, we will see what has caused Japanese scholars to misunderstand the B&L theory. To remedy this misunderstanding, I will introduce a recent study of politeness in discourse in Japanese (Usami 1999, 2002), which demonstrates how to analyse Japanese politeness in a quantitative manner, while at the same time considering how to treat *Keigo* not primarily as an indicator of politeness but as a way of satisfying the minimum requirement to be acceptable rather than *im*polite.

Given this background discussion, we will then see how vocal paralinguistic cues may contribute to conveying politeness. Here we build on studies of the speech act which reflect social/psychological distance and politeness. Most of these studies have associated politeness with prosodic features, such as F0 and duration. Even though some of these studies have implied that voice quality, especially breathiness, is involved in politeness, they have not explored this area yet. This is one of the main motivations for my study.

2.2 Universal perspectives on politeness (Brown and Levinson 1987)

To deal with politeness in language usage in a universal way, Brown and Levinson (1987) proposed their politeness theory. This theory is one of the most influential studies of politeness in linguistics, because of its "universality", its focus on language usage, and

also its quantitative manner of analysis to show the association between parameters and strategies used. I will introduce the details of this theory in this section, and I will discuss the applicability of this theory to speech analysis in later sections.

2.2.1 Model Person (MP): "Rationality" and "Face"

The B&L politeness theory assumes that every speaker and listener is a Model Person (MP), "who is a wilful fluent speaker of a natural language, further endowed with two special properties – rationality and face." In this specific context, they regard their MP as "rational" when s/he can define precisely the mode of reasoning which guarantees inference from ends or goals to the means that will satisfy those ends. In other words, the MP is rational enough to take one of several strategies to achieve her/his goal.
The other important concept in this theory is the notion of "face". By "face", they refer to two particular desires of the MP. One desire is to be unimpeded, and the other is to be approved of in certain specific respects. Brown and Levinson also suggested cross-cultural regularities of the MP in language usage, that is, that the MP uses linguistic strategies as a way of satisfying communicative and face-oriented goals, in a system of rational 'practical reasoning.'

2.2.2 The Concept of "Face": Positive and Negative

The concept of "face" represents the MP's *two desires*, "to be approved" and "to be unimpeded". Brown and Levinson defined the components of face related to the two desires as follows.

Positive face the desire of every member of society that his/her wants be desirable to at least some others: the positive consistent self-image or 'personality' claimed by interactants.

Negative face the desire of every 'competent adult member' of society that his/her action be unimpeded by others: the basic claim to territories, personal preserves, rights to non-distraction.

Brown and Levinson's argument for treating politeness in a universal perspective stands on this core concept of the relationship between "face" (= two desires of MPs, to be approved and not to be impeded) and "politeness" (= the way to achieve the desires of MPs, oriented toward the "face"). When they introduced their notion of two types of "face", they pointed out that "negative face" is familiar when associated with the "formal politeness" that is conjured up by the notion of "politeness", whereas positive face and its derivative forms of positive politeness are less obvious. This is because positive politeness is exchanged through actions which come from the mutual understanding and knowledge between interactants in a specific group. An example of positive politeness would be when a visitor comes to Mrs. A's garden, knowing that Mrs. A spends a lot of time and effort taking care of roses, and then praises her roses. Brown and Levinson emphasise that persons want to satisfy not only their desires, but to make sure that these desires are also desirable by some particular others especially relevant to these particular desires.
For example, Hearer (H) wants some persons (namely a_1, a_2, a_3 ....), such as:

a_1: set of all the classes of persons in H's social world.

a_2: set of all the persons in H's social stratum.

a_3: H's spouse.

to want the corresponding set of H's wants (w_1, w_2, w_3 ...), such as:

w_1: H has a beautiful front garden; H is responsible and law abiding.

w_2: H has a powerful motorbike and a leather jacket.

w_3: H is happy, healthy, wealthy, and wise.

Means to achieve these goals depend on both cultural and individual contexts, and therefore the use of positive face in a society is often highly restricted. For example, it would be unwise to assume that I (the author) am in the set of persons who would please you (the reader) by commenting on your clothes, because such a comment may cause affront unless I know you very well and commenting on appearance is acceptable to the community in which you and I belong. Brown and Levinson agree that this relationship is subject to rich cultural elaboration, as well as the core notions of "face" and "politeness" which are

universal among communities.

In this study, our focus is on Japanese politeness, which is believed to be based on negative politeness (as seen by the British, according to Brown and Levinson, 1987). We will revisit the argument about Japanese negative politeness later when we discuss *Keigo*, since this may help clear up the misunderstandings on the part of many Japanese scholars towards the B&L politeness theory.

2.2.3 Face Threatening Act (FTA)

Given the assumed universality of face and rationality of MP, we may imagine that certain kinds of acts intrinsically threaten face. In other words, these "acts" run contrary to the face desires of the hearer (H) and/or of the speaker (S). These acts are intended to be done by verbal or non-verbal communication. When Brown and Levinson defined Face Threatening Acts (FTAs), they classified FTAs according to two dimensions: firstly, what kind of face (positive/negative) is considered to be under threat, and secondly, whose face is threatened. In this classification, there is an overlap, because some FTAs intrinsically threaten both negative and positive face. The possible set of strategies for performing FTAs is schematised as Figure 2.1.

"On record" is a strategy which is taken in the situation when it is clear to participants that it is the communicative intention of a person to do an FTA. For example, if a person says "I (hereby) promise to come tomorrow" and if participants would concur that, in saying that, the person did unambiguously express the intention of committing himself to that future act, then in this terminology, the person went "on record" as promising to do so.

In contrast, if a speaker takes an "off record" strategy in doing an FTA, then there is more than one unambiguously attributable intention, so that the person cannot be held to have committed himself to one particular intention. For example, if a person says "Oh, I'm out of cash, I forgot to go to the bank today", the person may be intending to get someone to lend him cash, but he cannot be held to have committed himself to that intent. Linguistic realisation of off-record strategies include metaphor and irony, rhetorical questions, and understatement, which are all different types of hints as to

what a speaker wants or means to communicate, without doing so directly, so that the meaning is negotiable to some degree.

Doing an FTA "without redressive action, baldly" involves doing it in the most direct, clear and concise way possible (e.g., "Do X!" for requesting).

By "redressive action", Brown and Levinson mean action that "gives face" to the addressee. This action attempts to counteract the potential face damage of the FTA, or to indicate that no such face threat is intended or desired. This redressive action takes one of two forms, depending on which aspect of face (negative or positive) is being stressed. Positive politeness is oriented toward the positive face of H (hearer), which is approach-based. Negative politeness is oriented toward H's negative face, which is essentially avoidance-based and focuses on his want not to be impeded.

Given the set of strategies below, the more an act threatens S's or H's face, the more S will want to choose a higher-numbered strategy, because these strategies are increasingly less risky.

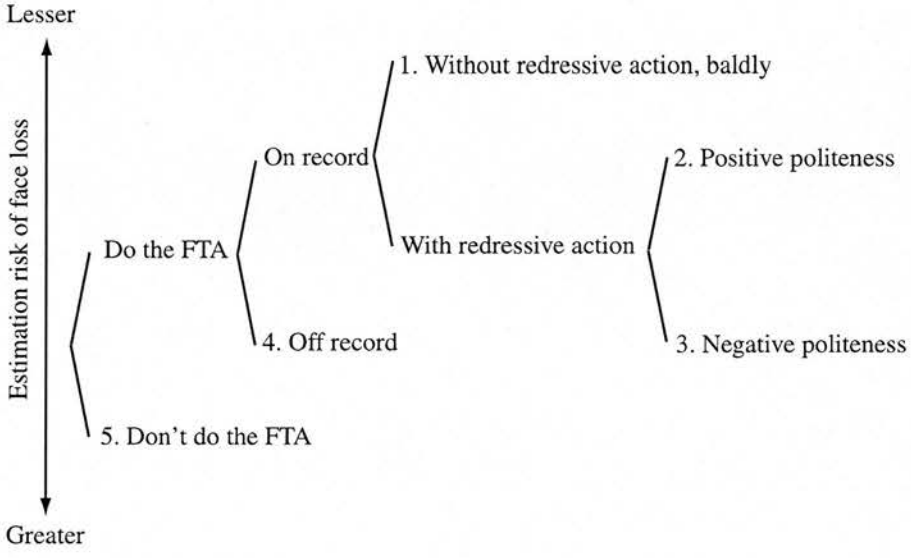


Figure 2.1: Circumstances determining choice of strategy for performing FTAs (From Brown and Levinson, 1987)

2.2.4 Degrees (norms) of politeness (Weightiness, Power, Distance and Ranking)

Brown and Levinson identified the following factors for assessing the seriousness of an FTA:

- the "social distance" (D) between S and H (symmetric)
- the relative "power" (P) of S and H (asymmetric)
- the absolute ranking (R) of impositions in the particular culture

Here, they take the view that P is a value attached not to individuals, but to roles or role-sets. They formulate the relationship between the weightiness of an FTA and the factors as follows.

$$W_x = D(S,H) + P(H,S) + R_x$$

W_x : the weightiness of the FTA $_x$
$D(S,H)$: the social distance between S and H
$P(H,S)$: the power that H has over S
R_x : the degree to which the FTA $_x$ is rated an imposition in the culture

2.2.5 Realisation of politeness strategies

Figure 2.2 is a summary of the classification of sixteen super-strategies of positive politeness, according to S's aim.

As Brown and Levinson state, positive politeness simply represents the normal linguistic behaviour between intimates. Positive-politeness utterances are used as a kind of metaphorical extension of intimacy or similarity, or as a kind of social accelerator, where S indicates that he/she wants to "come closer" to H. As an MP, S can assume that D (social distance) is short enough that s/he can employ positive politeness strategies for expressing their politeness.

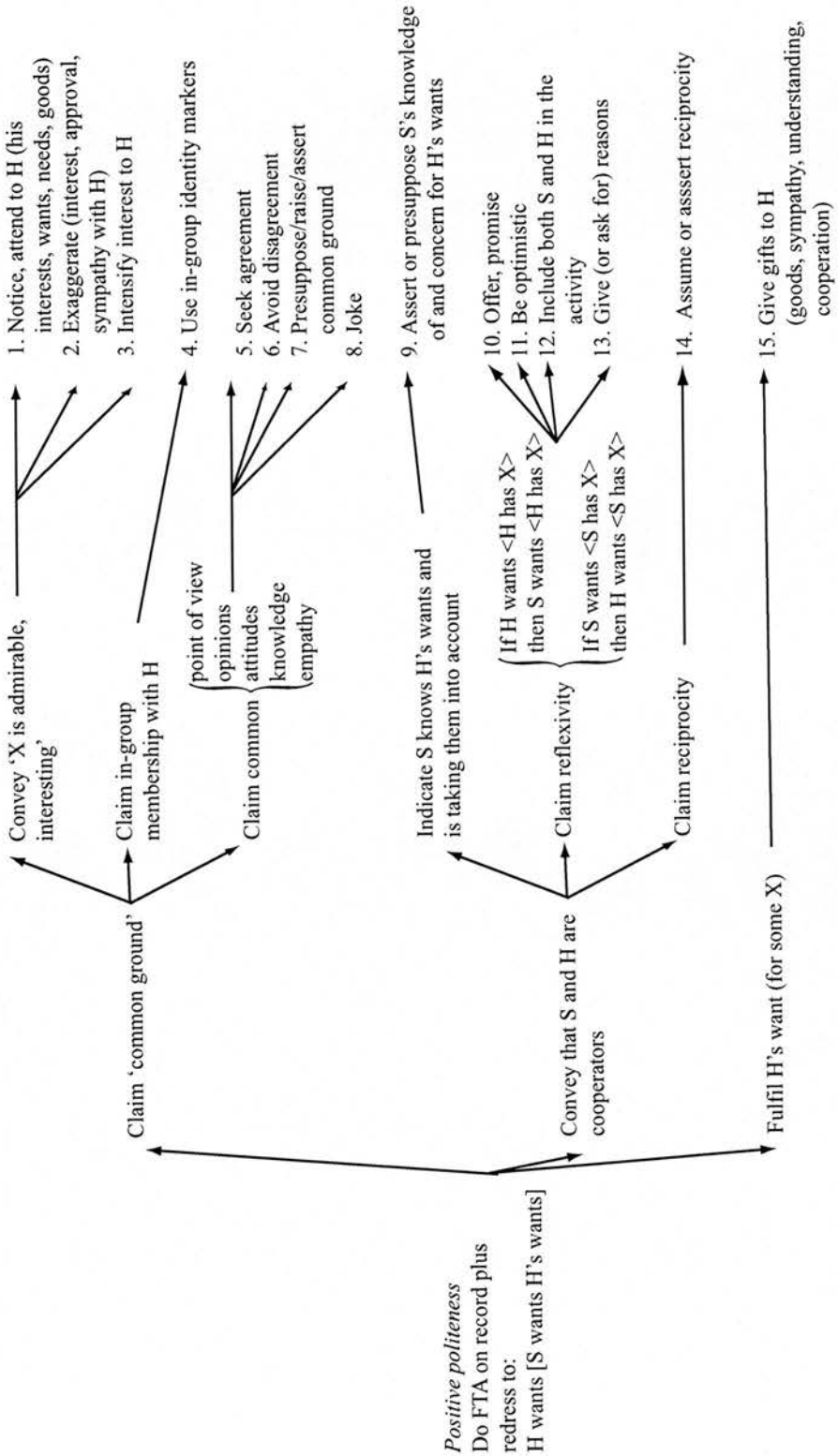


Figure 2.2: Super-strategy of Positive politeness (From Brown and Levinson, 1987)

Figure 2.3 is a summary of the classification of ten super-strategies of negative politeness, according to S's aim. According to Brown and Levinson, positive politeness is free-ranging, maximising intimacy or familiarity with joking and flattering in many ways, whereas negative politeness is specific and focused, and it performs the function of minimising the particular imposition that the FTA unavoidably creates. Brown and Levinson also stated that "When we think of politeness in Western cultures, it is negative-politeness behaviour that springs to mind," (Brown and Levinson 1987: pp.129-130). Most of what is treated in etiquette books is a matter of negative politeness.

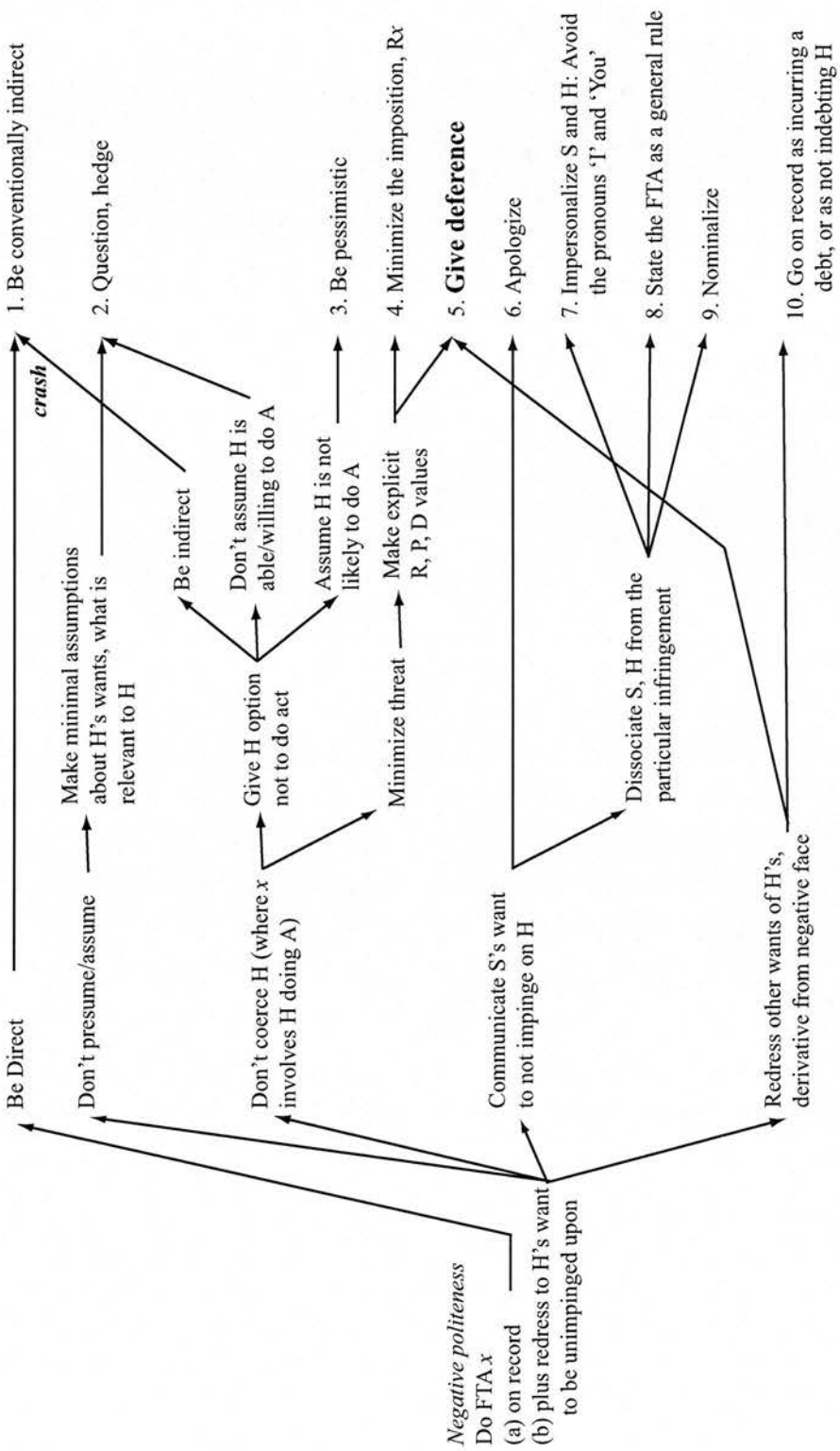


Figure 2.3: Super-strategy of Negative politeness (From Brown and Levinson, 1987)

2.2.6 *Strategies in Japanese from the Brown & Levinson politeness theory*

In the previous section, we overviewed the B&L politeness theory and the nature of the super-strategies. For some readers, quite a few strategies may not seem to belong to politeness, since the choice heavily depends on the culture and society they belong to. Therefore, we need to 1) understand the nature of the super-strategies, and how they may contribute to achieving politeness, and 2) consider the influence of the social/cultural background on the politeness strategy to be chosen in a particular community.

In the case of Japanese, the most likely strategy to be chosen is to "give deference", which belongs in the category of negative politeness. Within this strategy, there are two possible ways to realise deference. One way is for S to humble himself/herself and the other is for S to raise H. In both cases, the main message to be conveyed is that H is of superior status to S. Brown and Levinson analysed this case as follows. "Where reciprocal deference occurs, what is conveyed is a mutual respect based on a high D value, but this seems to be an exploitation of the asymmetrical use of deference to convey an asymmetrical social ranking."

Here, Brown and Levinson define the use of 'honorifics' as direct grammatical encodings of relative social status between participants, or between participants and persons or things referred to in the communicative event. In this connection they note the usage of plural pronouns to singular addressees in other languages. They also mention the effect of referent, bystanders and settings on honorific use. In the case of Japanese, they demonstrate how the system of honorifics ***Keigo*** functions in Japanese communication, including the awareness of all of these factors in addition to the relationship between speaker and addressee.

If we interpret Japanese "literally", from a very superficial point of view, *Keigo* functions as "giving deference" as Brown and Levinson stated. However, we need to examine Japanese linguistic behaviours more carefully so as to interpret what kind of politeness can be conveyed through *Keigo* in spoken communication. For this purpose, in the following section, I explain *Keigo* from both a systematic viewpoint and a viewpoint of practical usages. This will lead us to understand the politeness expected to be expressed in Japanese culture, further counter arguments set by Japanese scholars and will lead us to bring

remedies to the applicability of the B&L politeness theory by introducing the concept of marked/unmarked politeness, as suggested by Usami (2002).

2.3 *Keigo* and politeness, Face Threatening Act

2.3.1 Keigo: Reflection of Japanese politeness on linguistic structure

Here, I briefly summarise the function of *Keigo*, only as far as is relevant to politeness theory, and only as it affects the design of my study, since my aim is not to deal with the in-depth grammatical aspects of *Keigo*.

According to Shibatani (1990), there are two types of so-called "honorific processes" in Japanese, one along a speaker-addressee axis (formality axis) and the other along a speaker-referent axis (deference axis). In Japanese grammar the honorific system controlled by the speaker-addressee axis is called *teineigo* "polite-language" (e.g. verbal endings "-*de*su", "-*masu*"). That controlled by the speaker-referent axis is divided into *sonkeigo* "respect language" and *kenzyoogo* "humility/humbling language". Shibatani classifies *teineigo* as polite forms, and *sonkeigo* and *kenzyoogo* as honorific forms. In contrast to the speaker-addressee axis, honorific forms are correlated with deference, according to Shibatani (1990). He suggests distinguishing "formality" from "deference" according to the axis to be controlled.

One of the problems that Shibatani is attempting to deal with here is that *teineigo* is generally translated as "polite language" and is conventionally categorised as a part of the honorific system in Japanese grammar. This is partly because in English, "politeness" tends to be equated with "formality" and does not reflect the whole range of politeness dealt with in the B&L politeness theory. Shibatani's suggestion is that *teineigo*'s function is to express formality alone, and is not necessarily involved in "honorific" processes in the strict sense.

The relationship between speaker and addressee on the two axes proposed by Shibatani is considered along with another factor, namely whether the referent is in closer relationship to the speaker or to the addressee, shown in Figure 2.4 (as suggested by Brown and Levinson, 1987). This is another factor which affects the usage of *Keigo*. If a referent is strongly related to an addressee, this strong relationship is considered in the usage

of *Keigo* as well. For example, if a referent is a family member of the speaker and the addressee is not a member of the family, then the referent is referred to with *kenzyoogo*. On the other hand, if a referent is a family member of the addressee, and the addressee is a social superior of the speaker, then *sonkeigo* will be used to refer to the referent.

Shibatani stated that the speaker-referent axis and the speaker-addressee axis are independent. However, Inoue (1989) mentioned that the use of *teineigo* and honorifics (*sonkeigo and kenzyoogo*) is governed by a general requirement of concord or harmony over the use of honorifics and *teineigo*. This is because the use of the honorific axis depends on the relationship between a speaker and a referent, whereas the use of the formality axis depends on the relationship between a speaker and an addressee (and additionally on bystanders and the setting, according to Brown and Levinson, 1987). In daily conversation, the speaker tends to refer to the addressee. Therefore we seldom observe a situation controlled by the honorific axis, such as *sonkeigo* or *kenzyoogo* use, without the concomitant use of *teineigo*.

When a speaker is showing respect (controlling the speaker-referent axis), there is also necessarily an increase in formality (controlling the speaker-addressee axis). Here are some example sentences which correspond with the situation illustrated in Figure 2.4, where the speaker is talking to the addressee who is a teacher of the speaker and the referent is (1) a daughter of the addressee, (2) a mother of the speaker.

(TOP: topic marker, *SON*: *sonkeigo*, *KEN*: *kenzyoogo*, TEI: *teineigo*, prefix: pre, and suffix: suf, in the following examples)

(1) O-joo-san-wa o-dekake-ni natte i-masu.
pre(SON)-daughter-suf(TEI)-TOP pre(TEI)-go out do-verb**(SON)** be-suf(TEI)
"(Your) daughter has gone out."

(2) Haha-wa dekakete ori-masu.
mother-TOP go out be**(KEN)**-suf(TEI)
"(My) mother has gone out."

However, the usage of *teineigo* does not necessarily require the usage of *sonkeigo* or *kenzyoogo* as long as a referent does not need to be respected or self-humbled by a speaker.

Here is an example sentence (3) when the addressee is a stranger to the speaker, which can be compared with the example sentence (2).

(3) Haha-wa dekakete i-masu
mother-TOP go out be**(neutral)**-suf(TEI)
"(My) mother has gone out."

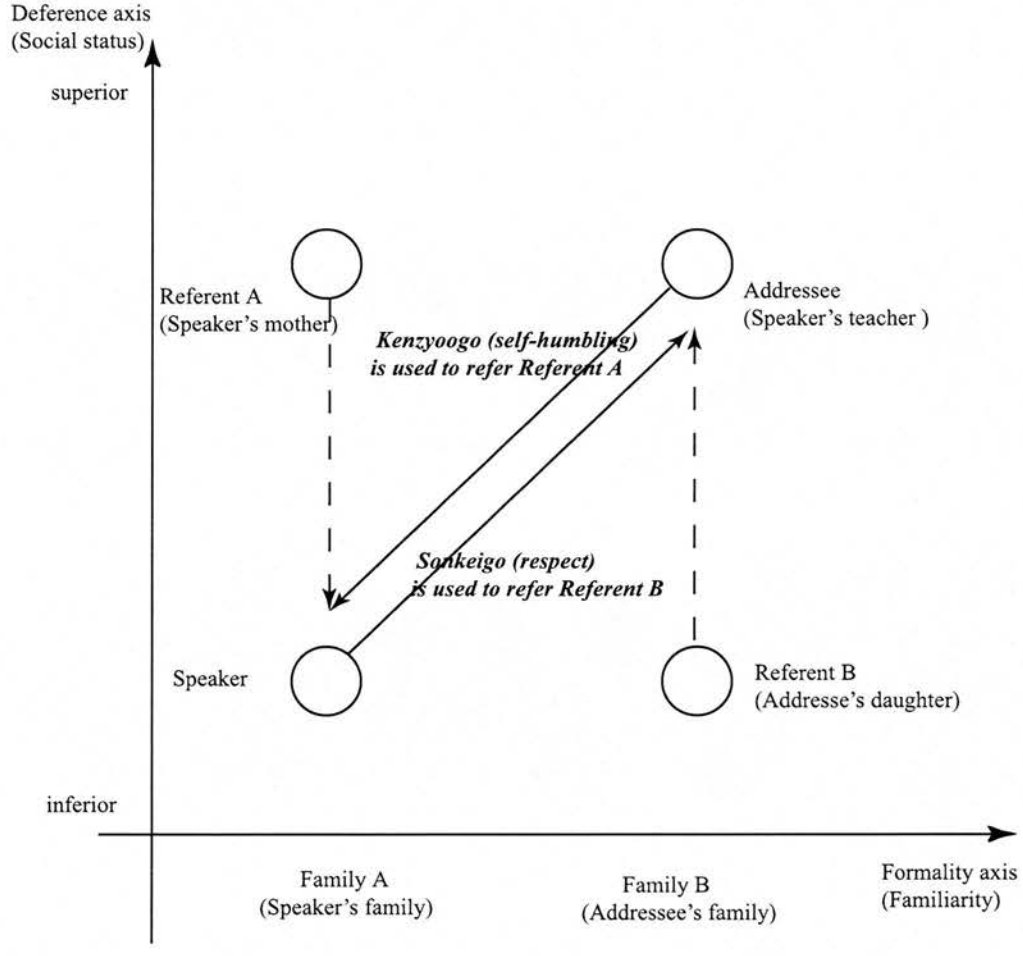


Figure 2.4: An example of schematic diagram of when a speaker changes the *Keigo* form according to the referent's attribute. Dashed lines show that the *Keigo* level shifts toward the referent, according to the familiarity with the speaker and the addressee. So lid lines' vertical directions show the *Keigo* level actually used by the speaker.

2.3.2 *Keigo: Interpretation and misinterpretation of Brown & Levinson politeness theory.*

If we interpret *Keigo* literally, it satisfies the definition of honorifics proposed by Brown and Levinson (1987). However, as pointed out by Shibatani, the actual usage of *Keigo* is for expressing "expected" formality and deference specified by R, P, and D, as if using grammar correctly. This could be interpreted as a part of politeness, but not always for expressing S's desire for "giving deference". *Keigo* includes the matter of formality according to the rigid relationship between people with different status or social distance, and does not always express only deference.

Shibatani (1990) also highlighted an interesting point that *Keigo* is affected by an insider-outsider distinction (refer to the example in Figure 2.4, because *Keigo* is demanded by social/psychological distance in Japanese society). Inoue (1989) mentioned that *Keigo*'s function is necessarily to keep a certain distance between people. Therefore, familiarity or psychological proximity with the addressee affects the usage of *Keigo*, in the sense that the use of *Keigo* toward someone who is unfamiliar is more likely to occur. Here we need to consider the notion of "*Ingin-Burei*", which is an expression meaning "very formal and rude". In a case where an addressee expects a sign of closeness from a speaker, yet the speaker uses a very formal form of *Keigo*, full of *sonkeigo* and *teineigo* towards the addressee, the addressee might well perceive the speaker's aim to keep distance from the addressee as rudeness. Usami (2002) cites an extreme case of *Ingin-Burei* in using *sonkeigo*. [1] When *sonkeigo* is used to an addressee in an intimate relationship, it could express the speaker's mounting anger. For example, if a wife says to her husband the following sentence (4), its direct translation is a factual statement, "you allow me to go home" with honorific suffixes.

(4) watakushi jikka-ni kaera-sete itadaki-masu
I(TEI) my parents' home-to go back-allow (SON(you))-suf(TEI)
"I'm allowed to go back my parents' home (by you)."

[1] Usami mentioned *sonkeigo* as "super-polite form", which is equivalent to "respect-form."

What she really means is "I do not want to be with you" and is suggesting separation or divorce. Usami shows in this case that choosing an inappropriate level of *Keigo* can be an FTA as well.

2.3.3 *Criticism of the Brown & Levinson politeness theory : misinterpretation of politeness*

Up to now, we observed the *Keigo* system that reflects the idea of Japanese politeness consisting of formality and deference, as accepted within Japanese society. On the basis of this idea, some Japanese scholars doubted the validity of the B&L politeness theory.

One of the best known counter arguments against the notion of "Western" politeness was suggested by Ide (1989). She conducted some perception experiments to compare a notion of "polite" in American English with a notion of "*teinei-na*" (adjective form of "*teinei*") in Japanese so as to show the contrasting perception of politeness. One of her interesting findings is that an adjective "friendly" is perceived as a component of "polite" by American people, whereas an equivalent adjective "*sitasige-na*" is not only perceived as a component of "*teinei-na*", but can be perceived as the opposite of "*teinei-na*" by Japanese people. Here, we see a conflict between the notion of "*sitasige-na*", which is accepted as one of politeness in Western society, and "*teinei-na*".

In fact, the usage of "*teineigo*", which is translated as "polite-form", can be perceived as "non-polite(neutral)" rather than "polite" in the system defined by Ide (1989). This is because it actually reflects the degree of formality in the conversational situation, so the use of "*teineigo*" should be taken as the degree of formality, as Shibatani (1990) mentioned. Usami (2002) also pointed out that proper usage of *Keigo* is taken for granted in speaking Japanese, and that poor usage of *Keigo* is regarded as "*im*polite". Brown and Levinson (1987) pointed out a strong similarity between "negative politeness" and "formal politeness". However, the appropriate usage of "*teineigo*" is just "formal-form". Again, *Keigo* reflects psychological and social distance between speakers, but not always

deference, even though this is what is literally expressed. Even the usage of "*sonkeigo* (respect-form)" and "*kenzyoogo* (humbling-form)" does not play the role of "giving deference", if this usage does not reflect properly the actual social status of the speaker and the addressee and psychological and social distance between them (P, D, and R in the B&L politeness theory). Therefore, the inappropriate usage of *Keigo* might express rudeness (Matsumoto 1988, Usami 2002), since both Japanese language and society are "social-context sensitive". The usage of *Keigo* may exaggerate positive politeness only in cases where it exaggerates P, D and R. A positive politeness strategy, "exaggerate", may be applied, but it runs a risk of demonstrating "*Ingin-Burei*" (e.g. suggestion of separation, or such as addressing a close friend with an unusual honorific title.)

Therefore I would like to conclude that the misuse of the word "polite" for the interpretation of *Keigo*-system ("polite-form" for "*teineigo*") has caused some confusion among Japanese scholars. First, "*Teinei-na*" is not equivalent to "polite", but rather just one of many forms of politeness which belongs to negative politeness, as defined by Brown and Levinson (1987). Second, the exaggerated usage of *Keigo* may express rudeness. The degree of deference expressed by *Keigo* does not necessarily correspond with the degree of politeness but rather with the desire of the speaker to keep distance from the addressee.

2.3.4 The concept of "Face": Another misunderstanding of definitions

The most important counter argument from Japanese scholars against the notion of "face" by Brown and Levinson (1987) is as follows. The notion of individuals and their rights has been acknowledged as playing an increasingly dominant role in European and American culture, however, this notion cannot be considered as basic to human relations in Japanese culture and society (Nakane 1967, 1972; Doi 1971, 1973; as quoted in Matsumoto 1988). Matsumoto quoted Nakane's opinion that the primary relations in Japanese society are between people who are related hierarchically in a certain social grouping rather than relations between people having the same status. Matsumoto illustrated this notion using the evidence that people introduce themselves not by their own personality (e.g. occupation, profession etc.), but by the society/community they belong (e.g. company, school etc.). Matsumoto also discussed Doi's argument, which

said that Japanese people prefer to be accepted by others rather than to insist on their individuality. Matsumoto concluded from this that the classification of *Keigo* as "giving deference" by Brown and Levinson is not appropriate, simply because the assumption of having "negative face" (that is, being unimpeded) does not fit with Japanese culture, as described by the Japanese scholars above.

However, there is a serious misunderstanding of the B&L politeness theory by these Japanese scholars. Brown and Levinson mention the possibility that the notion of face is linked to the ideas of the nature of the social persona: honour and virtue, shame and redemption. In the case of Japanese, the notion of face, which is used commonly, is highly associated only with honour and shame. However, Brown and Levinson were aware that what they defined as "face" in their theory is different from these notions. Brown and Levinson defined the notion of "face" to be valid only as the MP's physical property within the B&L politeness theory. Thus, it is clear that Brown and Levinson's aim is neither to generalise the notion of face specific to each culture, nor to force the notion of MP's face across cultural boundaries either.

It is true that Japanese people consider the social context seriously while producing utterances. But there could be several reasons for this: they may want to satisfy their desires "to be accepted as a member of society/community", or they do not want to be impeded in their goal "because of their improper usage of *Keigo*", or the rigid and hierarchical society of Japanese may take the improper usage of *Keigo* as *im*polite. In all these cases, they would still like to satisfy their "face" as MPs. Therefore we can say that these desires of a Japanese MP drive him/her to handle *Keigo* appropriately. Here we need to modify the classification of the usage of *Keigo*, as its previous definition includes both positive politeness and negative politeness. Using *Keigo* does not necessarily convey positive politeness, as believed by some Japanese scholars, because it is assumed that the Japanese MP will handle it properly. Nor does it necessarily belong to negative politeness, which is what we would infer from Brown and Levinson's classification of *Keigo* as a strategy of giving deference. Rather, *Keigo* is intended just to meet the state of the Japanese MP; using it is neither positive nor negative politeness.

Therefore, when we conduct experiments on Japanese politeness, the appropriateness of *Keigo* by participants should be examined so as to confirm that the participants satisfy the condition of the Japanese MP and therefore they perform tasks in a way expected by Japanese society. However, we should not decide whether an utterance is either "polite" or "impolite" based only on the proper usage of *Keigo*.

2.3.5 The Brown & Levinson politeness theory applied to Japanese

As discussed in the previous section, many Japanese scholars have questioned the adequacy of the B&L politeness theory, mainly because of the different interpretations between the linguistic systems of Western languages and Japanese. Considering these arguments, Usami (1999, 2002) proposed the perspective of "Discourse Politeness" to solve these problems. Overall, she supports the B&L politeness theory, and she regards the theory as still applicable with some appropriate modifications. Firstly, she points out that limiting the focus of analysis to utterance level is not appropriate since focusing on specific speech acts fails to incorporate the functions of discourse-level behaviour. Secondly, the B&L politeness theory is heavily inclined towards S's strategic language use, and does not consider the potential use of language as a means of conforming to social norms and conventions. Here, she suggested that politeness should be viewed from two angles, "language use that conforms to social norms and conventions", and "individual speaker's strategic language use". Jorden (1962) expressed the similar viewpoint as an American teacher of Japanese in the early 60's. The usage of *Keigo*, especially the level of formality and politeness (such as familiarity), is determined by the formality of the situation and of the individuals involved as well as by the familiarity of their friendship. In conversations between persons occupying different positions in the social scale (e.g. employer and employee, customer and salesperson), the person in the lower position uses more deference and a more formal level of speech, and this is taken for granted.

Thus, the important concept of "unmarked politeness" is proposed by Usami (2002). "Unmarked politeness" is characterised by avoiding the impoliteness caused by the improper usage of *Keigo*. Since this politeness is taken for granted in daily conversation,

this usage is rated as "non-polite (neutral)".

Usami estimates the correlation between the degree of politeness and the effect as shown below (Figure 2.5). Here she defines the risk of FTA according to the gap between the degree of politeness estimated by S and H.

Appropriateness of Action	Impolite	Appropriate	Excessively polite (*Ingin-burei*)
Gap between risk estimation (E) of FTA by S and H	-1=<Es-Eh<-α	-α=<Es-Eh<=+α	+α=<Es-Eh<=+1
Effect of politeness	Negative effect	Neutral/Positive effect	Negative effect

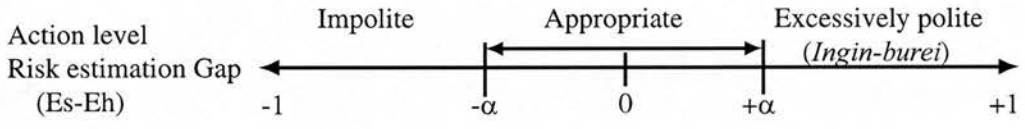


Figure 2.5: The degree of politeness and the effect of politeness (Usami 2002). The upper table corresponds with the risk estimation gap axis. α: threshold of absolute value of politeness, S: Speaker, H: Hearer, E: estimated degree of FTA, Es: E of Speaker, Eh: E of Hearer

The distinctive elements of Usami's analysis of politeness are 1) her introduction of the concept of "unmarked politeness (neutral)" and 2) the successful design of her experiments, with controlling parameters to elicit "positive politeness" strategies defined by Brown and Levinson. This included having participants who are strangers to each other involved in cooperative tasks. She controlled P, D and R of participants according to varied gender and age, with a common level of familiarity (that is, strangers) for all participants. Since her aim was to analyse literally expressed politeness in discourse in particular, and her main focus depends heavily on the contextual level. She claims that an analysis of the utterance level is not suitable.

Indeed, her results show that deference forms did not follow the principle which reflects social status of the speaker and the addressee, and the appropriate use of speech-level shifts between different *Keigo* forms (e.g. start omitting *teineigo* when participants get to know each other in a cooperative task, in an approachable manner) serves as an indicator of positive politeness as described in the B&L politeness theory. However, this can be

observed only through the discourse as long as it is based on lexical cues. Therefore, she concluded that the discussion about politeness needs to consider the constraint of *Keigo* use that encodes socio-cultural values, but at the same time, it is more important to consider discourse-level phenomena in order to examine pragmatic politeness. As long as one is studying politeness based on lexical cues, Usami suggested that politeness should be studied as an interaction between the constraint of the principle of *Keigo* (that is, social and cultural context sensitivity) and strategic language use such as speech-level shifts and topic management.

More evidence which supports this constraint of the principle of *Keigo* is suggested by Eckert and McConnell-Ginet (2003). They summarised the function of *Keigo* in the context of society and gender as follows: "Marking social status is tightly integrated into the grammar of Japanese, and showing respect or deference is a central component of so-called women's language in Japanese" (2003; p.164). They also commented on the situation of Japanese females within Japanese society as follows: "While all Japanese deploy honorifics, women's place in the social hierarchy constrains them to 'honor' others in their speech more than men." (2003; p.166) Eckert et al. agreed that Japanese people, both males and females can handle *Keigo* at a certain level as MPs. However, they also state that this social constraint of the *Keigo* use is applied especially to females due to the fact that *Keigo* expresses refinement and propriety and therefore this usage indexes femininity. Thus *Keigo* use is strongly associated with a gender role.

In summary, Usami states that analysis on the utterance level is insufficient to capture pragmatic politeness. Although this point is clearly valid, we appear to convey politeness in daily conversation at utterance level as well. Under certain circumstances, our utterances reflect the gap between the formality of *Keigo* presented on record, and the actual politeness which resulted from the overt signalling of S's desire. Even from one utterance, we can still sense some pragmatic degree of politeness, whether S is really polite, or simply "*Ingin-Burei*". For instance, when the usage of *Keigo* is very formal, and simultaneously the vocal expression does not show the same degree of politeness, as illustrated at the beginning of Chapter 1, what will listeners perceive? In this case we can tell that S is

impolite enough to express his/her desire to keep his/her distance from H, or try to do an FTA to H. This is the case of "negative face" of S, not to be impeded, according to the B&L politeness theory.

2.4 Politeness and speech

In previous sections, we have considered the interpretation of the B&L politeness theory, and have argued that it is applicable to Japanese with some appropriate modification, especially with regard to the understanding of *Keigo*. The key to applying the theory to Japanese is to adopt the interpretations of "face" and "*Keigo*" proposed by Usami (1999, 2002):

- We can use the notion of the MP who has "face" which meets social expectations in the Japanese cultural/social context.
- The strategies to be taken for the usage of *Keigo* are not only to "give deference", but also to "satisfy the strong social imposition (strong R) not to be impolite (unmarked politeness)".

Usami's ideas, however, apply particularly to the use of *Keigo*, and to pragmatic politeness expressed at the overall level of discourse. She does not consider the use of paralinguistic cues at all. It may be that, by considering such cues, we will find evidence for some kind of politeness expressed at the utterance level as well. Specifically, in Japanese, there may be situations in which we use more cues at utterance level, to signal that S wants to **shorten** the distance from H, without using the context of the discourse to express this greater closeness.

Such a situation may arise when participants from the same community or society are engaged in a cooperative task in an efficient way, and they want to express more positive politeness somehow to enable smoother communication between them in each utterance. Since *Keigo* is highly and rigidly dominated by hierarchy and social distance between speakers (superior and inferior, or strangers), especially in a way intended to keep a distance between them (Inoue 1989), there may be another channel to conveying positive politeness to shorten distance between speakers and to express friendliness or solidarity.

Ogino and Hong (1991) conducted a survey and many people responded that they took *Keigo* as a clue for deference whereas they took voice as a clue for familiarity, friendliness and solidarity. In this section, I introduce some examples that suggest that politeness of this sort can be conveyed by vocal paralinguistic channels, and discuss the appropriate use of these channels.

2.4.1 Politeness and speech: Evidence from Brown and Levinson (1987)

Brown and Levinson give some examples of phonological/prosodic usage for expressing politeness. Tzeltal speakers use creaky voice to express positive politeness and falsetto for negative politeness. In Tamil, people use high-pitched voice to signify deference from a low-status speaker to a high-status addressee when heavy FTA assessments were estimated. Brown and Levinson emphasised not simply that there are correlations of prosodic or phonological features with social contexts, but rather that there are rational reasons why these particular features are used in these particular circumstances. Creaky voice has a natural low speech energy and is therefore suitable to convey calmness and assurance, but not for negative politeness. High-pitch has a natural association with the voice quality of children, and if this feature is used to an adult speaker from an adult addressee, it may implicate self-humbling and thus deference. Brown and Levinson predicted that sustained high pitch would be a feature of negative-politeness usage and creaky voice a feature of positive-politeness usage, and it would not occur in a reverse manner in any culture.

2.4.2 Frequency code: Is pitch used to express politeness by Japanese male speakers?

As suggested by Brown and Levinson (1987), pitch is one of the suprasegmental features well known to signal politeness. Here I introduce the concept of the "frequency code" and some experimental studies to test the appropriateness of this concept. I will also briefly discuss this concept's applicability to speakers of Japanese.

Since pitch is a perceptual attribute which correlates with physical F0 (fundamental frequency), the quantitative analysis of F0 is important. Ohala (1984, 1996) explored the use of F0 in speech, where the sound-meaning correlation showed cross-language consistency. To extend the interpretation of the use of high pitch, he introduced the notion of the "frequency code". For example, low pitch/high pitch is equivalent to strong/weak or large/small. He went on to state that the sound-meaning correlations found in these cases adhere to his concept of "frequency code", which also governs the vocalisations of other species, where high F0 signifies smallness, non-threatening attitude, and desire for the goodwill of the receiver, whereas low F0 conveys largeness, threat, self-confidence, and self-sufficiency. He also mentioned the existence of sexual dimorphism in the vocal anatomy such that the male has a large larynx and a larger vocal tract not only in humans but in many other species as well. In spite of the fact that the evidence for the affective use of F0 is not as extensive as the use of F0 to mark sentence types, he concluded that "social" messages such as deference, politeness, submission, and lack of confidence are signalled by high and/or rising F0. Bolinger (1964, cited in Ohala 1984) suggested that messages such as assertiveness, authority, aggression, confidence and threat are conveyed by low and/or falling F0.

In the case of Japanese speech, there are several experimental studies about the use of the frequency code, based on quantitative analysis of F0 and the perception of physical and psychological attributes across several cultures.

Ogino and Hong (1991) analysed utterances produced by professional voice actors/actresses and amateur speakers. Ogino and Hong also confirmed that in the case of female speakers the pitch range of utterances with a high degree of perceived politeness is significantly higher (significance level<0.05) than that of utterances with a low degree of perceived politeness. No similar tendency was found in male speakers.

Bezooijen (1995) studied socio-cultural aspects of pitch usage between Japanese and Dutch female speakers. The results of her experiment showed three major findings:

1. The positive association of high pitch with attributes of physical and psychological powerlessness (short, weak, dependent and modest) was found in Japanese culture,

but not in Dutch culture.

2. There was a stronger differentiation between the ideal woman and man, in terms of powerlessness/power, in Japan than in the Netherlands
3. There was a preference for high pitch in women in Japan and for medium or low pitch in women in the Netherlands.

These results confirm the assumption that Japanese women raise their pitch in order to project a vocal image associated with feminine attributes of powerlessness. With regard to the first result, Bezooijen also found evidence for the universality of the frequency code suggested in Ohala's theory: when speaking at a high pitch, speakers from both cultures are perceived by listeners of both genders and from both cultures as shorter, weaker, more dependent, and more modest than when speaking at a low pitch. She concluded that Japanese women raise their pitch in order to conform to socio-cultural expectations stressing femininity. The Japanese ideal woman differs from that of Dutch mainly in the aspect of dependence, and no other scale showed significant differences. Bezooijen added that there also seemed to be strong expectations for the Japanese ideal man to be taller, stronger, more independent, and more arrogant. However, this is a purely anecdotal and impressionistic opinion suggested in her discussion and no hard evidence for male speakers was provided by this study.

Ohara (1999, 2001) also studied the socio-cultural aspects of pitch usage in a Western language and Japanese, based on the case of bilingual speakers of American English and Japanese. Here, the term 'bilingual' is used to emphasise the fact that they possess a high level of proficiency in their second language. The bilingual speakers consisted of two groups: Japanese-English bilinguals, who have Japanese as their first language and American English as their second language, and English-Japanese bilinguals, who have American English as their first language and Japanese as their second language. Ohara controlled the status of the addressee (professor/friend) when collecting speech from these bilinguals in Japanese and English. The results showed very interesting findings, such as:

1. Both Japanese and American females employed a higher F0 when speaking Japanese than when speaking English.

2. Japanese female speakers employed a higher F0 to superior addressees, only while speaking Japanese, whereas this tendency was not observed in American females.
3. Male bilinguals did not show any significant difference in F0 use either between languages (Japanese/English) or as a function of the status of addressee.

From these results, Ohara stated that this difference in F0 use within each speaker did not seem to be caused by linguistic structure, physiological features, or psychological state. For example, not only Japanese females but all the subjects were asked to convey the same messages to a different status of addresses in English and Japanese both, thus it is not a linguistic feature which affects F0 use according to the addressee's status. The Japanese female subjects used their own articulatory organs to produce the same messages to addressees of a different status, thus it is not a physiological feature of the speakers which affects F0 use according to the addressee's status. Also the messages were controlled in a way so as not to change the psychological state hugely. Taking evidence from her previous studies into account, Ohara asserted that Japanese society has a strong expectation for female members to employ a higher pitch to sound cute, soft, gentle, kind, polite, quiet, young, and beautiful whereas speakers of English in the United States are not under the same cultural constraints regarding the pitch of their voice. Furthermore, Ohara (2001) found that English-Japanese bilinguals are well aware of the social meanings attached to voice pitch in Japanese, and their responses confirmed that voice pitch is one of the key components for establishing female gender identity in Japanese society. There is evidence that female learners of Japanese as a second language also make an effort to realise this association between pitch usage and gender role when they accept the woman's role (after experiencing conflict over the fact that females are still treated unequally in Japan). Ohara therefore believes that informing female learners at beginning level about this correlation between voice pitch and gender identity in Japanese culture may be effective in assisting to demonstrate their competence in speaking natural Japanese expressing the socially expected degree of femininity.

Thus, we can conclude that the use of the "frequency code" suggested by Ohala seems applicable in a case where female speakers emphasise and demonstrate their femininity to

meet the desired expectation in Japanese society. On the other hand, this F0 use is unlikely to be seen in male speakers of Japanese, because it works in a way which disqualifies male speakers from the ideal male gender role. For this reason, the use of the "frequency code" for other purposes is prohibited for male speakers in normal conversation.

2.4.3 Speech rate: Universality and individuality

Speech rate is another suprasegmental feature which has been associated with politeness in Japanese. Here I introduce relatively recent studies with quantitative analysis of speech rate.

Ogino and Hong (1991) analysed utterances produced by professional voice actors/actresses and amateur speakers. They found that utterances produced with slower speech rate and longer pauses are likely to be perceived as "*teinei-na*" (refer to section 2.3.3, since the translation of this term is problematical). This correlation between tempo and a sort of politeness ("*teinei-na*") was found in both male and female speakers. Another study of the role of speech rate in perceived politeness was conducted by Hirose et al. (1997). They analysed utterances which were lexically similar, though not identical, and found that a "*zonzai-na*" (brusque) speaking style seemed to have faster speech rate than neutral speech.

In both these studies, the speech that was analysed was obtained by asking the participants to act, e.g. to pronounce a sentence politely. In observing the "acting voice", there might be 1) unnaturalness from acting, and 2) an interaction between politeness and other emotional factors.

Ofuka et al. (2000), in their study of utterance-level influences on perceived politeness in Japanese speech, observed the F0 movement, speech rate, and duration of the final morae of each utterance in their sample. They suggested that the final vowel of the utterance had a great impact on politeness judgements. They also concluded that the function relating politeness and speech rate was that of an inverted U-shape (Figure 2.6). The utterance sounds polite when speech rate is moderate, whereas it sounds more impolite when speech rate is increasingly faster or slower. However, they also stated

that the rating of politeness with speech rate is dependent on listeners' preference, and listeners were likely to rate higher politeness when the utterance has a speech rate close to that of their own. They concluded that polite utterances require that every influencing feature (F0 movement and duration) be kept within a certain range. However, this varies depending on politeness level, speaker, and listener.

The unclear result of Ofuka et al.'s study seems to be due to the fact that they did not control types of politeness according to the politeness strategies to be achieved. Rather, they simply controlled the intended addressees (a respectable gentleman, a young student, and a shabby drunk) so as to elicit the attitudes such as "polite", "casual" and "authoritative" respectively. They took it for granted that the politeness expressed in the utterance would be situation-dependent in the stated way. They did not seriously examine the likely strategy adopted by the speakers or relate their predictions to any overall picture of Japanese politeness. The only exception to this statement is that they considered a possible gender effect - the expression of femininity. The two main components of Japanese politeness, "formality" and "deference", were not even mentioned.

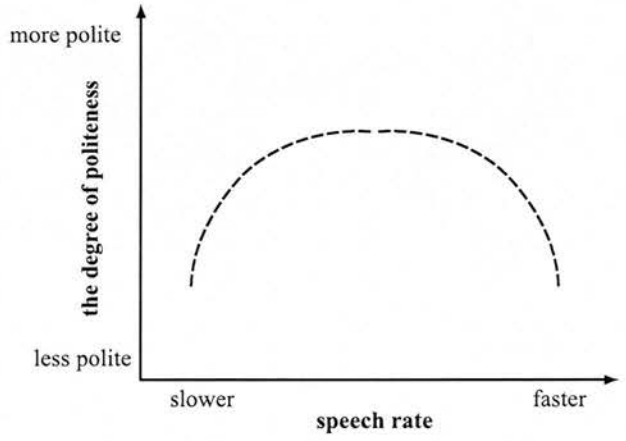


Figure 2.6: Correlation between speech rate and the degree of politeness: inverse U-shape relationship suggested by Ofuka et al. (2000)

It is worth noting that Laver (1994) predicted the findings by Ofuka et al. about speech rate and politeness as universal tendencies. The function of an inverted U-shape may be explained by the fact that the rate of the speaker may be correlated with the mood of a speaker. For instance, fast speech may express irritation or urgency, whereas a slower rate

may show hesitancy, doubt, or boredom in statements or sympathy or encouragement in questions and commands (Ramsaran 1989, quoted in Laver 1994). As for the listeners' preference that is suggested by Ofuka et al. (2000), Laver stated: "Only familiarity with a given speaker's habitual performance allows a listener to make an accurate assessment on any given occasion of the paralinguistic versus extralinguistic status of that speaker's patterns of continuity and rate of speech." (Laver 1994: pp.535) Thus, Laver mentions that the assessment of speech rate depends on the emotional and physical state and background of the speaker, and the listener's knowledge about the speaker helps this assessment, but this assessment does not depend on the formality of the situation.

2.4.4 Voice quality and paralinguistics

Another vocal paralinguistic feature mentioned by Brown and Levinson is voice quality. Laver (1991) comments that "paralinguistic features are used on a relatively ephemeral medium-term basis, while the same features, when used as phonetic components in voice quality, are quasi-permanent." In this study, I address voice quality which is sustained through an utterance, as opposed to that of the phonetic components, since the degree of politeness to be expressed does not change in a short term as the phonetic component does. Here I summarise paralinguistic features of voice quality relevant to politeness. The detailed physical characteristics of voice quality will be discussed in the next chapter.

Laver (1980) discusses some instances in which voice quality is involved in the signalling of politeness. Apart from the case of the high-pitch used for honorifics (Brown and Levinson 1978, cited in Laver 1980), he suggested the various kinds of sustained voice quality which contribute to conveying politeness as follows. Creaky voice signals bored resignation in English when it is used throughout the utterance. The use of whispery voice is wide-spread, and likely to be used as an indicator of secrecy and confidentiality in English and many other cultures. Breathy voice is not likely to be used phonologically as often as whispery voice; however, it is exploited in English for the communication of intimacy.

Another possibility is that breathiness may function as an alternative to F0 for signalling deference, because breathy voice is more likely to be observed in female voice. There have been some studies as follows. Moore(1939) states that breathy tone tends to be judged as lower in dominance and higher in introversion. Addington (1968) suggests that increased breathiness in male is associated with perception of immaturity, introversion, and self-doubt. Aronovitch (1976) associated a voice of less intensity and more breathiness with more submissiveness in males. Considering the fact that both female lexicon choices (Eckert and MacConnell-Ginet, 2003) and high F0 (Ohala, 1984, 1996) serve to signal deference, breathiness may well also serve to express deference without signalling femininity itself as obviously as do lexical choices and F0.

From the foregoing discussion, it seems likely that there are two different kinds of possible links between breathiness and politeness. First, breathiness may reduce psychological distance, as suggested Laver (positive politeness); second, breathiness may signal expressing deference and submissiveness (negative politeness). In using a breathy voice quality, a speaker may seek to achieve both types of politeness simultaneously.

2.5 Approaches to politeness studies in speech: Summary

2.5.1 Problems in previous politeness studies

In the previous section, I have reviewed earlier studies of politeness and speech, especially those which used 1) quantitative analysis of speech and politeness, and 2) impressionistic description of breathiness and politeness. In this section, I point out some problems these earlier studies encountered. Also I suggest some remedies which will be employed in my study to resolve these problems.

First of all, in view of the fact that Japanese society is still male-dominated, I believe it is important to study male speakers of Japanese as unmarked speakers. If we study female speakers only, which was done by Bezooijen (1995) and Ohara (2001), all we find

is marked politeness which is a feature of femininity. Outside the context of femininity, we may explore several politeness strategies, which may vary according to the situation. For example, Ohara (2000) analysed data collected from broadcast material. However, the data were taken from a programme designed for housewives who were educated to be "ideal woman", who are bored and killing time during lunch and teatime. The material consists of a considerable length of the female housewife audience's consultation about husband-wife relationships. The material has a theatrical effect and is controlled by the media's preference, which is to keep traditional roles of male and female. Therefore it could not avoid biased editing which would be required to make the show suitable for that kind of audience. The result was far from daily conversation in a neutral situation without the context of gender roles.

Ogino and Hong (1991) conducted a survey on what kind of politeness would be signalled by voice and they found that listeners did not detect the difference in the perceived degree of politeness in male and female speakers at utterance level. They studied the association of prosodic features such as F0 and speech rate and politeness. They found that high F0 use by females was the only difference on the production side of politeness between genders. Even though there have been some studies which show that female speakers of Japanese could signal their politeness by either highly skilled *Keigo* use (e.g. Eckert and MacConnell-Ginet 2003) or high F0 pitch as we have reviewed above, how male speakers would express their politeness is not fully explored.

A second issue is that the quantitative analyses stated above collected polite/non-polite utterances by asking speakers to act them out, and manipulated stimuli were used for perception experiments (Ogino and Hong 1991, Bezooijen 1995, Ohara 1999, Hirose et al. 1997, and Ofuka et al. 2000). Therefore naturalness of utterances was not guaranteed. This problem should be considered seriously to achieve reliable evaluation of expressive speech, as was pointed out by Campbell (2000). The exception to this is Ohara (2000), since she analysed data collected from spontaneous speech from broadcast material. However, as just noted, her material has the different problem of being seriously influenced by gender role factors. In my study, I employ a cooperative task,

which by its nature determines the participants' roles in the interaction. This task does not require any explicit direction of acting out, such as reading scripts, and thus we can obtain relatively natural utterances while avoiding acting effect.

A third issue is that the use of voice quality has not been seriously considered, mainly for the following three reasons.

First, most politeness studies have been based on lexical analysis of discourse as opposed to acoustic analysis.

Second, even in those studies that have been based on acoustic analysis, prosodic features such as F0 and speech rate were regarded as more likely to convey politeness, and therefore most studies of polite speech have been based on the "frequency code" and on speech rate.

Third, measurement methods for F0 and speech rate are relatively well established, whereas measurement of voice quality requires either complicated computation, such as inverse-filtering, or a painful and unnatural situation forced on speakers, and therefore it is difficult to obtain reliable acoustic features with paralinguistic information by either approach. As a result of the problems regarding voice quality measurement, Ofuka et al. (2000) suggested that breathiness might have an effect on politeness strategies influenced by Japanese culture and society, but presented no quantitative evidence. There have been some experimental studies of the communicative role of perceived breathiness, associated with males' introversion and submissiveness (Moore 1939, Addington 1968, Aronovitch 1976), but in an impressionistic approach only. In fact, no study of the communicative role of breathiness has taken a quantitative approach to acoustic analysis. Yet current techniques of spectral measurement allow us to easily measure a glottal configuration related to the breathiness of an individual voice, so this methodological shortcoming should be remediable. Discussion of the acoustic analysis of voice quality is the topic of the next chapter.

2.5.2 *Toward an optimal design for eliciting politeness in natural speech*

In summary, the goal of this study is to shed light on breathiness and how it may be used to express and perceive politeness in Japanese. The study is based on the following theoretical assumptions and methodological strategies, which have been argued for in this chapter.

Firstly, I treat *Keigo* use as "non-polite (neutral)" since it is taken for granted that the Japanese MP uses *Keigo* appropriately. This is especially true for people who are highly educated (Inoue 1989). I therefore recruited speakers who are competent in this regard, that is, mature and well educated enough to handle *Keigo* properly.

Secondly, I recruited speakers of superior and inferior ranks from the same community. Under this condition, formality based on the psychological distance between the participants was limited by familiarity, and thus expressions of deference should be due primarily to rank differences. The participants were set a cooperative task in order to elicit positive politeness as suggested by Usami (1999, 2002). This may allow us to observe whether their vocal expression matches the negative politeness such as deference from the social rank gap expressed by *Keigo*, or whether it cuts across to show positive politeness conveying solidarity and willingness to cooperate due to an effect of the task, or both types of politeness.

Thirdly, I recruited male speakers for the production test since my aim is not to study femininity, which includes not only politeness but also other features of Japanese society. Because my aim is to focus on politeness, "unmarked" speakers of Japanese were employed.

Fourthly, by manipulating the appropriate situation of the cooperative task, I was able to induce the speakers to produce spontaneous speech but within a limited range of vocabulary and intentions. This gave relatively controlled material for analysis while avoiding unnaturalness due to acting.

Finally, in order to observe breathiness, I used acoustic measurements of spectral and waveform parameters representing voice quality, especially breathiness, not on invasive physiological measurements. These acoustic measures are the subject of the next chapter.

Chapter 3

Breathiness: Physiological and acoustic aspects

3.1 Introduction

In this chapter, I summarise the theoretical and methodological approaches towards the study of perceived breathiness found in the literature.

Firstly, I introduce breathiness as a perceptual phenomenon which may contribute to conveying linguistic, paralinguistic, or extralinguistic information, via acoustic characteristics.
Secondly, I discuss physiological views on producing breathy voice. Here I also summarise some correlates of perceived breathiness with physical observations.
Thirdly, I discuss techniques to measure aerodynamic and acoustic parameters, as well as glottal configurations which correlate with perceived breathiness observed above. These techniques employ either acoustic measurements or equipment such as the Rothenberg Mask. Also the advantages and disadvantages of using these techniques will be discussed.
Finally, I summarise previous findings, which motivate the measurement method used for this study. The method primarily involves the direct waveform and spectrum measurement suggested by Klatt and Klatt (1990) and Hanson (1995). The validity of this measurement method is discussed briefly by comparing it with other methodologies that

use inverse filtering, such as direct aerodynamic measurement and glottal flow estimation from microphone recordings.

3.2 Voice quality and types of information it conveys

In our daily life, acoustic characteristics correlated to perception of voice quality, together with other prosodic features, may signal many kinds of information (Table 3.1).

Table 3.1: Types of information which could be conveyed by perceived voice quality (Fujisaki 1997, modified by the author) (Emotion is excluded from examples of information, because of controversy of controllability by speaker which results in difficulty of categorising it as paralinguistics or extralinguistics (Bard, personal communication)).

Type of information	Examples of information	Controllability by speaker	Time range	Speaker type
Linguistic	phonemic contrast	controllable	short-term	normal
Paralinguistic	**attitude**	**controllable**	**mid-term**	**normal**
Extralinguistic	speaker characteristics (age, gender)	normally uncontrollable (excl. voice actors)	long-term	normal
	disorder	uncontrollable	long-term	disordered

Linguistic information includes phonemic contrasts such as whether aspiration is found in consonants or in voice quality (e.g. Ladefoged and Antoñanzas-Barroso 1985, Gobl and Ní Chasaide 1988). Laver (1980) mentioned, however, that this phonemic contrast would be treated as short-term voice quality changes, which belong to the field of articulatory phonetics, and this voice quality change does not last more than a few segments.

In this study, our interest is in conveying politeness through voice, which falls in the category of paralinguistics. This sort of voice quality lasts in a mid term durational range (e.g. utterances). This type of information depends on the attitude and/or mood of a speaker and therefore it lasts more than a few segments, but does not last semi-permanently.

As for the voice quality in the long term, pathological studies and speaker characteristics studies used in forensic and speech technology applications have been conducted. Above all, many studies (e.g. Kreiman and Gerratt 1998, 2000; Shrivastav 2003) which explored perceived breathiness in the area of pathological voice, in which voice quality lasts permanently, have brought us a lot of informative results. These studies suggest to us the difficulty of objective measurement of perceived breathiness, associated with glottal and/or acoustic characteristics, and give us important information about how to overcome the known problems, such as difficulty in defining voice quality and obtaining listener agreement in voice quality scaling of complex auditory stimuli.

In this study, our target is a normal speaker's vocal expression. Therefore the speaker's voice variation is supposed to fall within the normal range of the glottal configurations so that measurement techniques used for examining normal speakers' voice would be appropriate while measurements exploited for pathological purposes may be less useful.

3.3 Glottal characteristics in the production of breathy voice

In this section, I introduce the different manners of voice production believed to result in breathy voice. Perceived breathiness is often associated with incomplete glottal closure. According to Titze (1994), breathy voice occurs when the average airflow is excessive at the glottis, which results in significant component of noise due to turbulence in the glottis.

For the aid of broad definitions of voice quality category terms, we can consult Laver's description of "phonatory settings". Laver (1980, 2000) described modal voice (a neutral, normal setting of the voice), harsh voice, creaky voice, falsetto voice, whispery voice, and breathy voice as shown in Table 3.2. These terms are especially applicable in a relatively long domain, such as an utterance, rather than in short time window (e.g. the voice quality of a single vowel).

Table 3.2: Phonatory settings and their characteristics (From Laver, 2000)

Phonatory settings	Characteristics
Modal	The regular periodic vibration of the true vocal folds Efficient air use in the vibration of the vocal folds Moderate muscle tension
Harsh	Dysperiodicity, with frequency jitter or intensity shimmer
Creaky	Low-frequency, irregular pulses
Whispery	Audible friction caused by air leaking through a slightly open glottis
Breathy	Lax muscular tension of all the phonatory muscle systems which lead to much greater air wastage than in the case of whispery voice
Falsetto	High pitch range with increased longitudinal muscular tension of the vocal folds and thin fold-edges

From the table above, breathy voice appears to be defined as involving some degree of whispery voice. Laver (1980) agreed that it is reasonable to acknowledge a close auditory relationship between breathy voice and whispery voice, because both involve the presence of audible friction. He suggested that the transition from breathiness to whisperiness is part of an auditory continuum, and that the placing of the borderline between the two categories is merely an operational decision.

However, the physiological relationship between breathiness and whisperiness is more distant than the auditory continuum as suggested by Laver (1980). Ladefoged (2001) and Catford (2001) stated that there are two kinds of breathy voice (refer to Figure 3.1): 1) the vocal folds vibrate loosely, remaining apart, so they appear to be flapping in the airflow (the same configuration as "voiceless" fricatives) and 2) the vocal folds are apart between the arytenoid cartilages in the posterior part of the glottis, where they can still vibrate, at the same time a great deal of airflow passes out through the posterior chink. The latter case is categorised as "whispery voice". Therefore, the mechanism of voice production with breathiness should be handled with care.

Hammarberg et al. (1986) emphasised that breathy voice as defined by Laver is a *breathy* and *hypofunctional (lax)* voice whereas whispery voice is a *breathy* and *hyperfunctional*

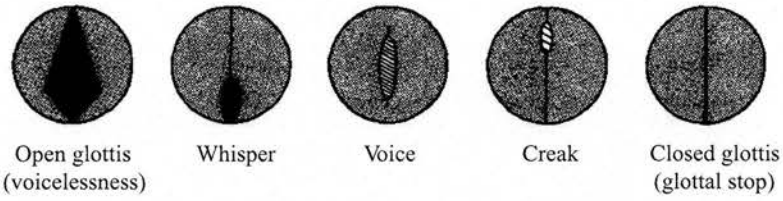


Figure 3.1: States of the glottis according to phonatory settings (From Catford 2001)

(tense) voice, more likely to be found in disordered voice. Here, hyperfunction comes from compressed vocal folds which results in strained voice, whereas hypofunctional voice has too little tension at the vocal folds. Hammarberg et al. defined breathiness as audible escape of air through the glottis due to insufficient closure.

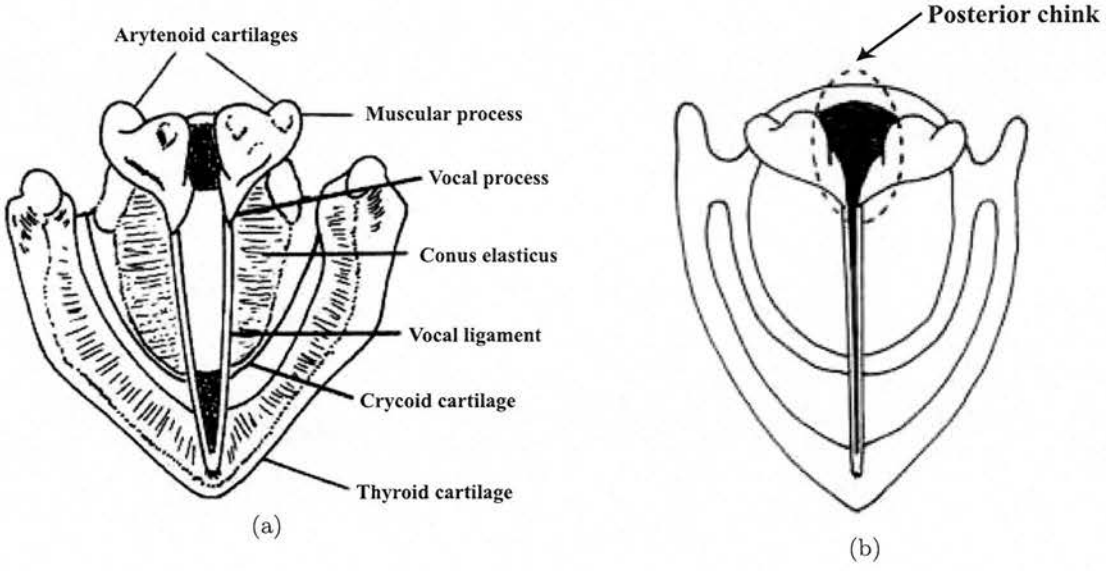


Figure 3.2: Illustrations of vocal folds (a) Superior view of the larynx illustrating the relation of ligament to arytenoid, crycoid and thyroid cartilages (b) focus on a posterior chink which is found as incomplete glottal closure in normal speakers. (From Kent 1997)

Titze (1994) stated that breathy voice results when the average airflow is excessive, and a pressed voice results when there is insufficient airflow. However, voicing actually ceases in extreme cases of both of these parameters, because vibration cannot be sustained when the vocal folds are either too widely separated or too tightly approximated. In pressed voice the vocal folds are hyperadducted, whereas in breathy voice they are hypoadducted (Figure 3.1). Titze mentioned that the air leakage at the glottis could be caused differently according to the speaker types. In normal speakers, the air leakage

normally occurs at the arytenoid cartilages (Figure 3.2) which is maintained during a phonatory cycle. On the other hand, the air leakage may occur at any other place along the vocal folds in the case of a disordered speaker.

Södersten et al. (1990, 1991) evaluated the correlation between perceived breathiness together with the degree of hypofunction and hyperfunction and the degree of incomplete closure at the glottis by fiberscopic observation of normal speakers across gender. Together with earlier clinical studies, they suggested that 1) clinical studies mainly focused on the membranous portion, whereas normal speakers tend to have a posterior glottal gap (chink) as incomplete closure of glottis, 2) females are found to have a more incomplete closure of the posterior parts of glottis, 3) the degree of perceived breathiness was judged to be higher for the female voices compared to the male ones, and most importantly, 4) it is a common tendency across genders that the degree of incomplete closure corresponds to the degree of breathiness. Södersten et al.'s findings suggest that the correlation between perceived breathiness and the degree of incomplete glottal closure, resulting from the size of the posterior glottal chink, could be applied to male speakers.

3.4 Perceived breathiness and its glottal configuration correlated with acoustic/aerodynamic measures

In this section, I summarise previous studies of the aerodynamic and/or acoustic characteristics which are found to be correlated with perceived breathiness. Further discussion of measurement methods will be discussed in the next section.

3.4.1 Aerodynamic measures relevant to perceived breathiness

3.4.1.1 Overview of aerodynamic measures

Given that an incomplete closure of the vocal folds leads to air leakage at the glottis, and that the degree of the incomplete closure corresponds to the degree of breathiness, we may obtain some insight into the degree of breathiness by observing glottal flow. A typical glottal flow waveform and its derivative are illustrated in Figure 3.3. In this figure,

the pitch period is indicated by T (1/F0), and the rise time and fall time of the glottal waveform are indicated by t1 and t2 respectively. The open quotient (OQ) is a ratio of the time that the glottis is open (t1+t2) to the total duration of the cycle of vocal fold vibration (T), which is defined as (t1+t2)/T in Figure 3.3. The skewness of the waveform, the speed quotient (SQ), is defined as the ratio of the rise time to the fall time, t1/t2.

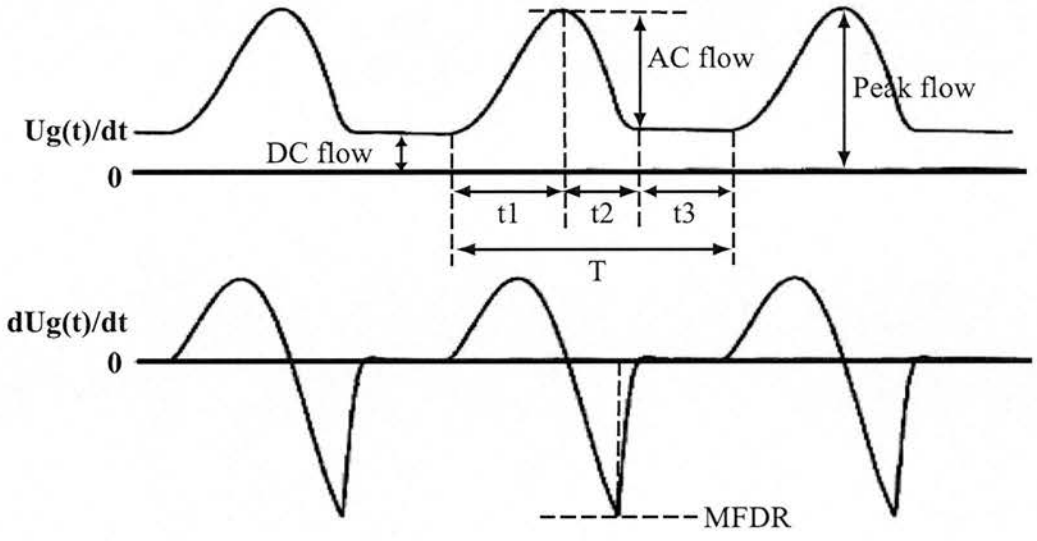


Figure 3.3: Glottal flow waveform, its derivative, and aerodynamic parameters (From Hanson 1995, Holmberg et al. 1988)

Holmberg et al. (1994a, 1994b, 1995) measured the aerodynamic parameters as follows (also refer to Figure 3.3):

1. Subglottal air pressure (= Sound pressure level (SPL))
2. AC flow (the modulated component of the glottal airflow wave form, reflects the magnitude of the vocal fold oscillation)
3. DC flow (minimum flow, which is the unmodulated component of the glottal airflow waveform)
4. Maximum flow declination rate (MFDR) which reflects the vocal fold closing velocity to produce the main excitation of the vocal tract
5. Adduction quotient (AQ), the ratio of the duration of the vocal fold closure in the duration of the glottal period (t3/T)

Titze (1993) stated that OQ values lower than 0.4 are often associated with a "pressed" quality and OQ values above 0.7 tend to have a "breathy" quality, especially at low pitches, because the DC flow is raised in relation to the AC flow.

Hammarberg et al. (1986) found that a *breathy* and *hypofunctional(lax)* (="breathy") voice is produced with low laryngeal effort, efficient glottal airflow, and increased DC flow. In contrast, a *breathy* and *hyperfunctional(tense)* (="whispery") voice is produced with high laryngeal and aerodynamic effort but lack of efficient glottal airflow, and the waveform is more likely to be skewed to the right and have increased MFDR and AC flow together with increased flow of DC flow.

3.4.1.2 Measurement method of aerodynamic parameters (Inverse filtering)

The aim of inverse filtering is to cancel the effect of the formant filter in the vocal tract so that we can estimate and observe source signals at the vocal folds, based on the source-filter theory of speech production. This theory takes the pair of vocal folds as the source of sound and the vocal tract as a complex of filters, the effect of which gives formant frequencies. Therefore, if the effects of the formants are removed from a speech waveform with a filter that is the inverse filter of the vocal tract, the glottal waveform will be extracted.

There are two approaches to the inverse filtering method. One approach is to measure direct oral airflow during speech production using a Rothenberg mask (Rothenberg 1973, cited in Hanson 1995) and estimate glottal flow after inverse filtering. The other method is to collect the speech waveform with microphone recordings, and after inverse filtering, to fit the derivative of estimated glottal flow to the models, e.g. the LF model (Fant et al. 1985, 1988). This method successfully preserves the information in the mid to high frequency regions of the glottal waveform. Most researchers have extracted parameters such as OQ directly from the glottal waveform that results from inverse filtering, or from a glottal waveform model that is fitted to the natural glottal waveform.

However, there are problems with regard to the accuracy of estimating formant location and bandwidth, necessary for cancelling out the filtering effect of vocal tract. Fritzell et al. (1986) reported that a slight mistuning of the first formant location could result in a different degree of air leakage, and manual estimation of formant frequencies and their bandwidths requires high skill, experience and a lot of time. Hertegård et al. (1992) reported that the extra pole-zero pair which appears in nasalised vowel leads to inaccurate estimation of formants. These sorts of inaccuracy of estimation of formants (vocal tract filter) can result in difficulties in decomposing a source signal from a complete speech signal.

3.4.1.3 Direct measurement of airflow

The Rothenberg Mask (Rothenberg 1973, cited in Hanson 1995) has an advantage in that it measures oral airflow directly from the mouth and so the mask can preserve a zero flow level (a measurement reference point recorded before each token is recorded) and avoid background noise interference which microphone recordings usually have. However, this mask has a limited bandwidth and it can process speech signal only up to about 1.1 kHz. To avoid the effects of a resonance in the mask just above 1.1 kHz (Badin et al. 1990, cited in Holmberg et al. 1995), the airflow needs to be low-pass filtered at about this frequency. Aspiration noise, one of the characteristics of breathy voice, lies in a higher frequency region and cannot be studied with this method.

Furthermore, there is a major disadvantage to all methods which employ equipment attached to a speaker's articulatory organs— not only to measure oral airflow, but also to measure glottal airflow (e.g. transducer (Cranen and Boves 1985, 1988)), or to observe the vocal folds (e.g. endoscopic system, fiberscopic system (Södersten et al. 1990, 1991)). All these methods restrict the free and natural movement of articulatory organs, and some of them may force speakers to produce speech in a very unnatural way. Therefore, such methods may be unsuitable for eliciting natural spontaneous speech, particularly if we are focusing on how speech conveys information about speakers' attitudes.

3.4.1.4 Microphone recordings: Estimation of glottal flow

Microphone recordings provide a non-invasive approach to extracting relevant parameters from the spectrum of the glottal waveform, or from the speech waveform and spectrum. Combined with microphone recordings, the most widely-used model to estimate glottal flow is the one suggested by Fant et al. (1985), the so-called LF-model (named after Liljencrants and Fant). In the LF-model (see also Equations 3.1, 3.2, and Figure 3.4), a glottal source is expressed with four independent parameters, summarised by Epstein (2002).

OQ **(Open quotient):** OQ is the ratio of the opening phase of the glottal waveform to the period of the glottal pulse, computed as follows. $OQ = (t_p - t_0)/(t_c - t_0)$.

R_k **(Glottal skew):** R_k measures the relationship between the closing phase and opening phase of the derivative of the glottal pulse, which is the inverse of the speed quotient. R_k is computed as follows. $R_k = \{(t_e - t_0) - (t_p - t_0)\}/(t_p - t_0)$.

E_e **(Excitation strength):** E_e is measured as the amplitude of the negative peak of the derivative of the glottal flow, and is the time constant of an exponential recovery in Equation 3.2 and Figure 3.4)

R_a **(Dynamic leakage):** R_a is a measure of the residual flow (return phase) from excitation to complete closure (or maximum closure, if there is a DC leakage). This measure is computed as the ratio of the time constant of the return phase to the period of the glottal pulse. $R_a = t_a/(t_c - t_0)$.

The main advantage of the LF-model is that it enables us to observe the behaviour of incomplete closure, or the residual phase of closure which is in progress after the discontinuity of the derivative of the glottal airflow. Therefore, this model is suitable for observing phenomena in breathy phonation.

t_0: the start of the glottal pulse

t_c: the duration of the entire pulse cycle

t_p: the duration of the time when the $dU_g/dt > 0$

t_e: the time when dU_g/dt reached the negative peak ($=E_e$)

E_e: the negative peak value of dU_g/dt

T_a: the effective duration of the return phase

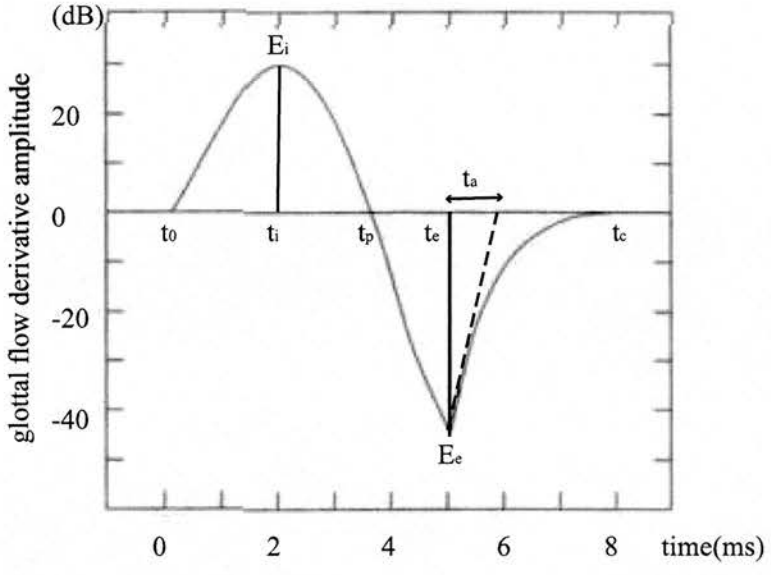


Figure 3.4: The LF-model of the derivative of glottal flow. The four wave-shaped parameters t_p, t_e, t_a and E_e uniquely determine the pulse ($t_c = T_0 = \frac{1}{F_0}$) (from Fant et al., 1985)

$$E(t) = E_0 e^{\alpha t} \sin \omega_g t \qquad (t_0 \leq t \leq t_e) \tag{3.1}$$

$$E(t) = -\frac{-E_e}{\epsilon t_a} \cdot [e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}] \qquad (t_e \leq t \leq t_c) \tag{3.2}$$

However, this microphone recording method requires very strict conditions, using a phase-true microphone. In addition, inverse-filtered microphone recordings are very sensitive to low-frequency noise, and therefore may pick up ambient noise. As a result, the absolute transglottal airflow cannot be measured from these recordings (Hertegård et al. 1992).

Also, the inverse filtering stage requires recursive analysis by synthesis steps in finding filters to cancel out the effect of formants, and there is a difficulty in obtaining the appropriate filter to fit the vocal tract characteristics. Similarly, the LF modelling also requires source matching process of manual interactive analysis so as to optimise source estimation, which is also recursive. However, those who have employed the LF-model have successfully correlated acoustic measures with estimated glottal flow. This has contributed to revealing the nature of voice quality, especially in the case of breathy phonation.

3.4.2 Acoustic measures relevant to perceived breathiness: Direct speech spectrum and waveform measurement

As stated above, glottal flow estimation using the inverse filtering method has enabled experimenters to observe possible correlations between physiological behaviour and perceived breathiness. However, this method cannot avoid certain problems: The use of the Rothenberg mask or fiberscope is an invasive procedure, and microphone recordings are sensitive to recording conditions, especially the response in the low frequency region, which may cause changes in the skew of the glottal flow waveform and the spectrum, not present in the original speech signal. Furthermore, the response in the mid to high frequency region might be neglected, which is important if we are interested in observing aspiration noise. To overcome these problems, some researchers have attempted to measure acoustic parameters assumed to be correlated with perceived breathiness by observing the waveform and spectrum directly.

3.4.2.1 Overview of acoustic measures

Titze (1993) stated that the overall tendency of breathiness is likely to have a large spectral slope with non-harmonic components between the harmonic lines (aspiration noise) in the high frequency region (refer to Figure 3.5, by Stevens et al., 1995). Normal vocal quality has a spectral slope of approximately -12dB/octave, whereas "breathy" quality, which Titze defines as the combination of "fluty" with turbulent noise (aspiration), has a large spectral slope. In addition, the spectrum is filled with non-harmonic components (noise) between the harmonic lines.

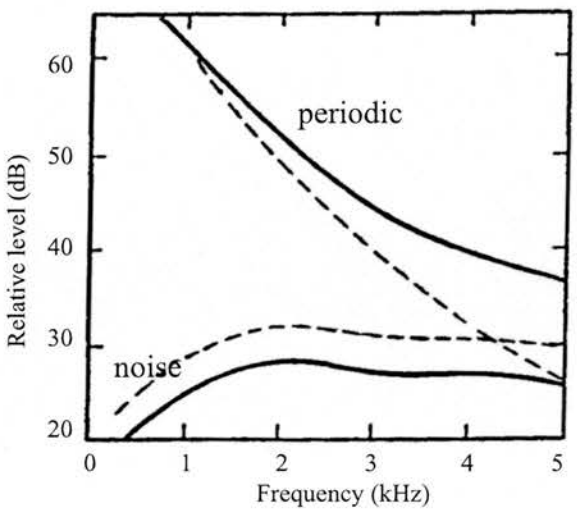


Figure 3.5: Calculated spectra and relative amplitude of periodic source and noise source for two different glottal configurations. Solid line: modal configuration, dashed line: more breathy configuration of glottal opening of 0.1 cm^2) (From Stevens and Hanson, 1995)

In early studies, Ladefoged and Antoñanzas-Barroso (1985) conducted experiments so as to correlate acoustics and production/perception of breathiness, comparing modal and breathy vowels. They computed two acoustic correlates, both spectral-based. Firstly, they computed the difference between the amplitudes of the first harmonic ($H1$) and the amplitude of the first formant ($A1$). Secondly, they computed the difference between the amplitudes of the first and second harmonic ($H2$). The results are as follows. The first correlate ($H1-A1$) shows a difference between modal and breathy voice within each speaker, but ($H1-A1$) is a relative amplitude and highly dependent on the individual speaker. A value which is modal for one speaker can be breathy for another speaker. Absolute values of the second correlate ($H1-H2$) seem to be more reliable in distinguishing modal and breathy vowels. However, when Ladefoged and Antoñanzas-Barroso conducted the breathiness perception test with native speakers of American English, ($H1-A1$) showed a high correlation with perceived breathiness.

3.4.2.2 Measurement method of spectral/waveform information

Klatt and Klatt (1990) studied various kinds of voice quality from pressed voice to breathy voice using natural reiterant speech to improve their formant synthesizer. Their method-

ology was also employed by Hanson (1995, 1997, 1999) who conducted measurements of female voices with acoustic and physiological measures related to perceived breathiness. Klatt and Klatt's method is distinct from others in that they relied entirely on acoustic measures extracted directly from the speech spectrum and waveform. By applying the compensation to the measures based on the speaker characteristics and vowel formant locations, we can possibly avoid recursive processes of source-filter decomposition with inverse filtering and LF modelling.

The following sections show the acoustic parameters of the speech spectrum and waveform employed in the analysis by Klatt and Klatt (1990).

$H1 - H2$ ***(a possible indicator of open quotient)*** $H1 - H2$ is the relative strength of the first harmonic, since it is known to increase together with the open quotient, which in turn is expected to be relevant to breathiness. However, Klatt and Klatt (1990) also discussed the difficulty of measuring the amplitude of the first harmonic ($H1$). Firstly, background noise with strong energy in the low frequency region could mask the detection of $H1$. Secondly, psychological equal-loudness contours indicate some attenuation of low frequencies relative to the first formant (Robinson et al. 1956, cited in Klatt and Klatt 1990). To determine whether the $H1$ is large or small, it needs to be compared with some reference which takes into account the recording level. Klatt and Klatt employed the second-harmonic amplitude ($H2$) as a reference in their analysis as suggested by Bickley (1982) (cited in Klatt et al. 1990).

Klatt and Klatt measured the $H1 - H2$ in the middle of each syllable position. For males, they found a tendency to slight laryngealisation during the f0 fall of utterance-final syllables which causes the reduction of the OQ. This tendency was found to be slightly more likely to occur in males. For females, they found that a speaker who is perceived to have breathy voice and presumably uses a speaking mode with a large OQ has a large value of $H1 - H2$, whereas another speaker who is perceived to have a laryngealised voice quality has a small value of $H1 - H2$.

From these results, Klatt and Klatt suggested that $H1 - H2$ might be used as an indirect measure of OQ.

Noise ratings (a method of quantifying possible indicators of aspiration noise) A method of noise ratings is suggested to quantify the presence of aspiration noise in the vowel spectrum, especially in the higher frequency region where the aspiration noise may replace harmonic excitation of the third and higher formants. Klatt and Klatt (1990) demonstrated this analysis in the following three steps. Firstly, the frequency of the third formant ($F3$) was estimated visually from a wideband spectrogram. Secondly, a four-pole Butterworth band-pass filter, having a centre frequency of $F3$ and a bandwidth of 600Hz[1], was used to create a filtered version of the original digitised waveform. Thirdly, a plot of the filtered waveform was rated visually using a four-step scale to determine the degree of random noise present. If the filtered $F3$ waveform consisted of a periodic damped sinusoid, in synchrony with the unfiltered waveform, the vowel was judged to be periodic and free from aspiration noise. If there was no visible periodicity in synchrony with the original waveform, then the vowel was judged to have strong aspiration noise. Hanson (1995, 1997, 1999) also employed this noise rating technique to estimate aspiration noise (Figure 3.6), and confirmed this method as valid since it showed a good correlation between raters.

Figure 3.6: The waveforms of synthetic tokens used for noise ratings by Hanson (1995), generated by using a bandpass-filter (center: 3kHz, bandwidth: 600Hz). The waveform of a token (left) with periodic component of moderate spectral tilt and aspiration noise. The waveform of the token (right) shows a token with larger spectral tilt in the source and aspiration noise. (from Hanson, 1995)

Tracheal coupling When the glottis is partially abducted, there may be extra resonances (poles) to be observed called "tracheal formants" due to acoustic coupling to the trachea. Klatt and Klatt (1990) found that the extra tracheal poles often distort vowel spectra to different degrees. When the tracheal poles are present, they tend to be located at frequencies consistent with the tracheal pole locations observed by other researchers

[1]The rationale of the selection of this bandwidth was not shown.

(Ishizaka et al. 1976, Cranen and Boves 1987; both cited by Klatt and Klatt 1990). However, Klatt and Klatt mentioned the difficulty of quantifying the coupling of the tracheal and lung system below the glottis objectively because of some potential perturbations present in their data.

*$BW1$ **(the bandwidth of first formant)** and $H1 - A1$ **(a possible indicator of** $BW1$**)*** $BW1$ is the width of the $F1$ peak and a cue for the relative strength or prominence of the $F1$ peak. The partially opened glottis of a breathy vowel may cause $BW1$ to increase (Klatt and Klatt, 1990). Klatt and Klatt measured $BW1$ using 1) the subjective visibility of $F1$ as a distinct local spectral maximum and 2) the relative amplitude of the first and second formant frequency. An alternative to this procedure for obtaining the local spectral maximum was suggested by Hanson (1995, 1997). Hanson estimated the relative amplitude of the $F1$ peak in the speech spectrum by measuring the amplitude of $F1$ relative to that of the first harmonic ($H1 - A1$).

Hanson (1995) pointed out that formant bandwidths are related to the rate of acoustic energy loss in the vocal tract. In the case of $BW1$, the energy losses in the frequency range of the first formant result from several factors, including the resistance of the yielding walls of the vocal tract as well as heat conduction and frictional losses at the walls. With airflow through the open glottis, glottal resistance can contribute to further energy loss, especially in the lower frequency region, meaning that the open condition of the glottis increases $BW1$. Therefore, measuring $BW1$ can provide an indirect indication of the degree to which the glottis fails to close completely during a cycle of glottal vibration.

Figure 3.7: The decay of $F1$ oscillation, used to estimate $BW1$. (From Hanson, 1999)
N.B. this figure shows an original speech waveform, not a bandpass-filtered waveform

Hanson suggested estimating $BW1$ from the speech waveform after bandpass-filtering around $F1$ (refer to Equations 3.3- 3.5, Figures 3.7, 3.8). If the $F1$ oscillation is assumed to be a damped sinusoid,

$$e^{-\alpha t}\cos 2\pi f t \qquad (where\ f = F1) \tag{3.3}$$

then the constant α (in per second) is related to $BW1$ by the equation,

$$BW1 = \alpha/\pi \tag{3.4}$$

Therefore, it is possible to estimate $BW1$ by measuring the decay rate of the first-formant waveform after bandpass-filtering around $F1$, during the early part of the glottal period (Figures 3.7, 3.8, Equations 3.4, 3.5).

$$BW1 = \frac{1}{\pi}\frac{\ln\left(\frac{x1}{x2}\right)}{\frac{1}{2}(t1+t2)} \tag{3.5}$$

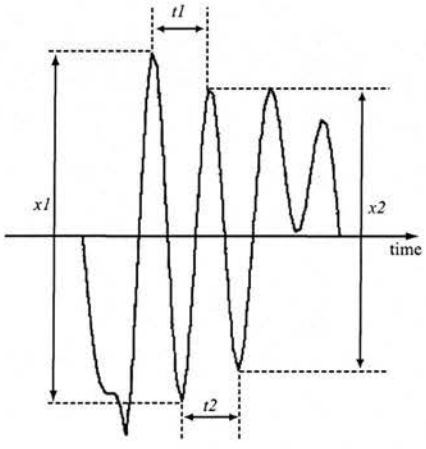


Figure 3.8: Stable oscillations for computation of $BW1$. (From Hanson, 1995)

where $x1$ and $x2$ represent the amplitudes of the peak-to-peak oscillations, and $t1$ and $t2$ represent the time between maximum and minimum amplitude of the oscillations (see Figure 3.8).

However, this method requires stability in oscillations, which is not always the case in male speech (refer to Chapter 4, the second glottal excitation of male speakers).

$H1 - A3$ ***(spectral tilt)*** Hanson (1995) suggested measuring the relative amplitude of the first harmonic and the third formant peak ($H1 - A3$) as a parameter of spectral tilt. A steep spectral tilt is most observable between the lower frequency region where F0 lies and the higher frequency region where the third formant peak lies. When the derivative of the glottal waveform has a discontinuity at the time of closing, the spectrum of the derivative at middle and high frequencies has a downward slope of 6 dB/octave at middle and high frequencies (see Figure 3.9). This spectrum is influenced by the abruptness with which the flow is cut off when the membranous part of the vocal folds closes during the vibration cycle.

For a given OQ, this abruptness can be affected in two ways, when there is complete closure of the glottis during some part of the vibratory cycle. One mechanism that leads to a change in abruptness is a glottal closing that does not occur simultaneously at all points along the anterior-posterior length of the vocal folds. Closing is a type of "zipper" action, with initial closure at the anterior end of the glottis and the closure sliding back along the length of the glottis. This closure leads to a more gradual cut-off of flow, resulting in a derivative of the glottal waveform that does not have a discontinuity. The effect on the spectrum is to introduce an additional downward tilt in the spectrum at high frequencies. From this discussion, Hanson conducted the following approximation. If T_D is defined as the duration of the zipper action, from the time of initiation of the anterior closure to the time of closure at the posterior end, and assuming that the gradual cut-off is exponential, then an approximation to the time constant T of this exponential is roughly one-half of the time of the sliding closure, that is,

$$T \approx \frac{T_D}{2} \tag{3.6}$$

The point where the slope of spectra changes is then given by

$$f_T = \frac{1}{2\pi T} = \frac{1}{\pi T_D} \tag{3.7}$$

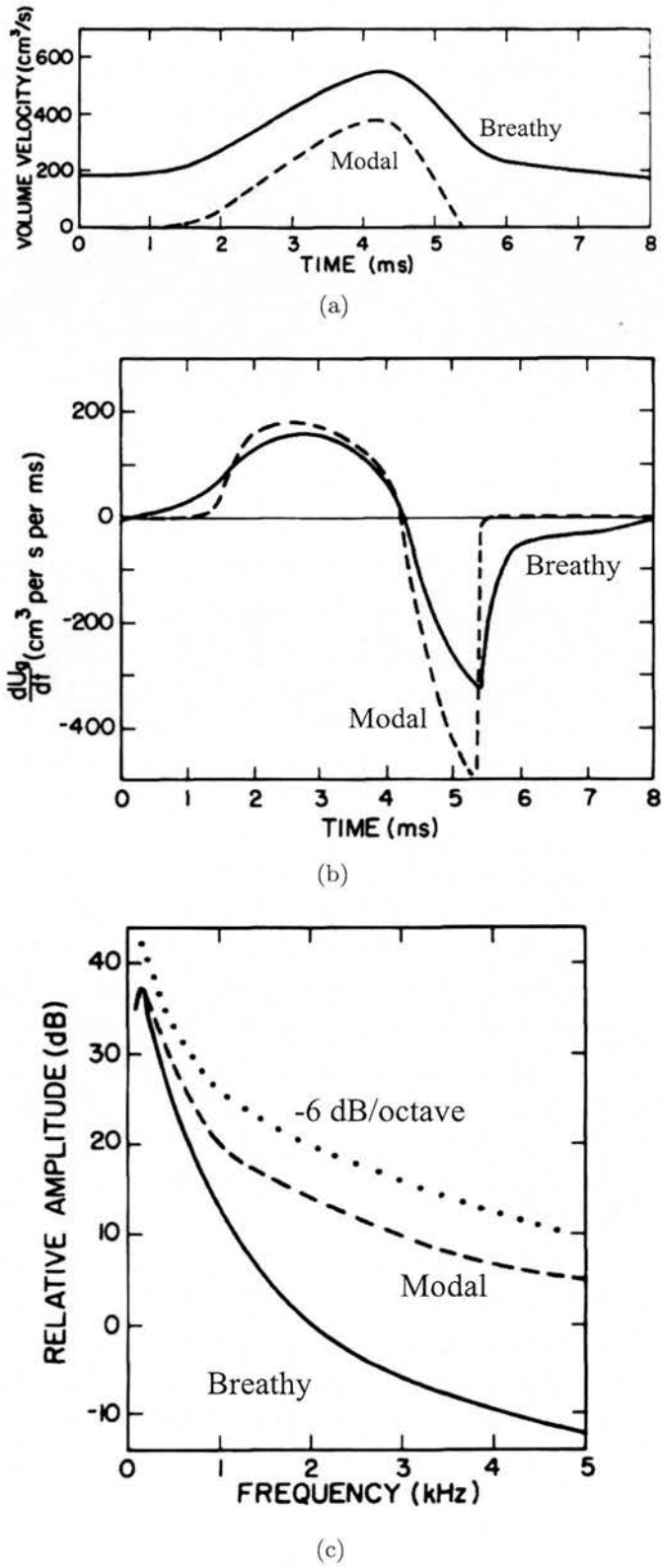


Figure 3.9: Comparison of airflows between modal and breathy voice (a) the glottal flows (b) the derivative curves of the glottal flows and (c) the spectral slopes of the derivatives modal and breathy voice normal speakers. (From Stevens (1997))

Above f_T, the slope of the spectra of the derivative of the waveform increases to 12 dB/octave if an exponential approximation is assumed. For $f_T$less than about 2 kHz, the resulting increase in the tilt at f_{avgf3}, an average location of $F3$ (for female speakers, 2750Hz) is

$$20 \log_{10} \frac{f_{avgf3}}{f_T} \tag{3.8}$$

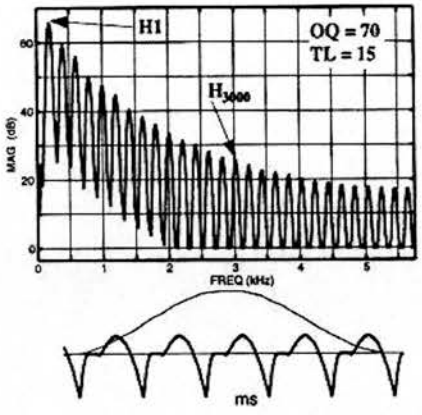


Figure 3.10: Spectrum and derivative of the glottal flow of the periodic glottal volume-velocity source, under the condition of the gradual cutoff flow of the vocal fold closure, generated by the KLSYN88 synthesizer (Klatt and Klatt 1990). This diagram is taken from Hanson (1995).

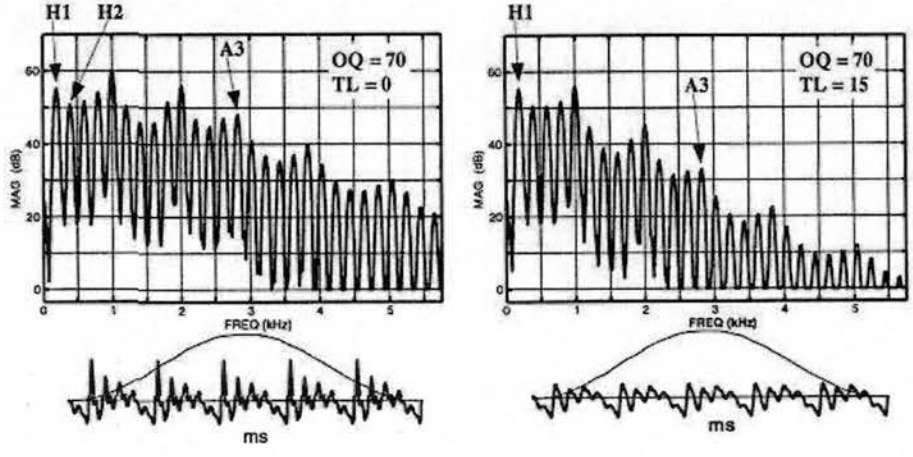


Figure 3.11: Spectrum and waveform (the [æ] vowel)of the periodic glottal volume-velocity source, under the condition of the same OQ but different TL, generated by the KLSYN88 synthesizer (Klatt and Klatt 1990). The diagrams are taken from Hanson (1995).

Hanson compared the spectra of tokens synthesised using the KLGLOTT88 model (Klatt and Klatt 1990) between the two configurations, where OQ are the same but TL (spectral

tilt) are different (Figure 3.11, also refer to Figure 3.6). She noted that the amplitude of the third formant ($A3$) dropped when TL increased. Thus she concluded that the amplitude of the third formant relative to that of the first harmonic ($H1 - A3$) is likely to be a reasonable and accurate indication of source spectral tilt, assuming that $H1$ is a good reference point to compare to $A3$. However, there are two issues to be considered: 1) $A3$ is likely to be influenced by the locations of $F1$ and $F2$, therefore, if comparing $A3$ across vowels or speakers, a certain correction needs to be made for this effect, 2) when comparing $A3$ across vowels, the radiation characteristic of each vowel affects the bandwidth of $F3$ to a greater extent than those of lower formants do.

3.5 Related work: aerodynamic/acoustic parameters

3.5.1 Physiological instrumental studies and acoustic parameters

Södersten et al. (1990, 1991) observed the degree of incomplete glottal closure with the fiberscope, and compared their observations with both perceived breathiness and the relative amplitude of the first harmonic and the first formant ($H1 - A1$) as suggested by Ladefoged et al. (1985). Like Ladefoged et al., Södersten et al. found that $H1 - A1$ is related to perceived breathiness, in addition to their findings on the degree of incomplete closure (refer to Section 3.3).

Holmberg et al. (1995) conducted a comparison between aerodynamic, electroglottographic and acoustic measurements of the female voice. Holmberg et al. measured sound pressure level (SPL), adduction quotient (AQ) (defined as ($t3/T$) in Figure 3.3), DC flow, and the $F3$ spectral component. As acoustic correlates, Holmberg et al. measured sound pressure level (SPL), the difference of amplitude between the first and second harmonics ($H1 - H2$), and fundamental frequency of glottal vibration (F0). They used the Rothenberg Mask, electroglottography (EGG), and direct recordings from microphone for extracting the parameters above. For the measurement of the glottal airflow waveform, they used the Rothenberg Mask. MFDR was measured from the first derivative of the glottal airflow waveform. For the measurement of AQ, they also employed EGG, in order

to complement the glottal airflow measures. To compensate for the effect of wearing the mask, SPL measured from the microphone recordings was adjusted so as to correspond with the level measured by the Rothenberg Mask. Apart from the problem of employing the Rothenberg Mask stated above, EGG measurement is also invasive. However, both these invasive methods also contributed to correlating the measurements of oral airflow with acoustic correlates, such as SPL and $(H1 - H2)$. Holmberg et al. suggested that the amplitude difference between $F1$ and $F3$ $(A1 - A3)$ was informative about the abruptness of the airflow decrease. They also suggested that $(H1 - H2)$ may be used as a substitute measure for flow-adduction quotient, when unsuccessful inverse filtering makes AQ measurements unreliable. The comparative studies of aerodynamic and acoustic correlates demonstrated that spectral parameters such as $H1 - H2$, $H1 - A1$, and $A1 - A3$ can be informative about glottal configurations, and can be obtained in a non-invasive manner.

3.5.2 Glottal flow estimation with the LF model: Analysis and Synthesis

Ní Chasaide and Gobl (1989, 1992, 1997, 1999) employed the LF model with inverse filtering for their voice quality analysis in comparison with other acoustic correlates, such as harmonics and formants, in order to clarify the nature of the phonatory settings defined by Laver (1980). Their comparison between different categories of voice quality puts emphasis on the following dimensions: Tense vs. Lax and Whispery vs. Breathy. The spectrum of tense voice is highly dominated by $F1$, while lax voice has the opposite tendency. The relatively higher amplitude of $F1$ in tense voice could be due to smaller losses which are reflected in narrower formant bandwidths. Breathy voice and whispery voice both have a similar tendency to lax voice, but in these two voice qualities, a greater attenuation of $F1$ and $F2$ is likely to be observed. Also, these two voice qualities are highly dominated by the first harmonic. In a recent study of perception and acoustic correlates, Gobl et al. (1999) suggested that spectral tilt appears to be a major determinant of perceived breathy voice. Gobl et al. suggested that other parameters such as the bandwidth of $F1$ and $F2$, or OQ (Open Quotient) and SQ (Speed Quotient of KLSYN88a by Klatt and Klatt (1990), derived from R_k of the LF-model by Fant et al. (1985)), alone cannot take a role in perceived breathiness. However, when these four

parameters are set to that of modal voice, the increase of spectral tilt does not appear to be a cue for perceived breathy voice by itself. From this result, even though spectral tilt is dominant in perceived breathiness, without the support of other parameters, we cannot determine that spectral tilt indicates the degree of breathiness.

Karlsson (1991) and Carlson et al. (1991) studied the glottal waveform of female speakers using the LF model, with the aim of improving female voice synthesis. Using GLOVE, the KTH speech synthesis system, they found that the control of the NA parameter, which simulates pitch-synchronous noise at the glottis, improved the performance of the synthesizer bringing it closer to the human voice.

From the studies above, spectral tilt and noise at the glottis are more likely to be cues for breathiness. However, none of these parameters can stand alone as a cue for breathiness.

3.5.3 Spectrum-based measurements

Klatt and Klatt (1990) conducted a perception test on the acoustic correlates of perceived breathiness. They found a contradictory relationship to that which was suggested in previous studies, between the "noise-in-F3" measure and the $H1$ measure. In an utterance-final syllable, when they observed more noise, indicating a greater glottal airflow, they also observed a weaker $H1$, indicative of pressed voice with a shorter OQ. They concluded that this contradictory relationship is caused by the tendency to open the larynx in preparation for breathing, resulting in an increase of posterior glottal chink and an increase in aspiration noise. From this result, they presume that most speakers rotate the anterior tips of the arytenoid cartilages inward to maintain voicing, and at the same time, partially laryngealise, which leads to a breathy-laryngealised mode of vibration. Another effect of an open glottis on the transfer function of the vocal tract is an increase in the bandwidth of the first formant ($BW1$). This increase can be quite large for a low vowel. Also, nasalization can have a similar flattening effect on $F1$ due to energy loss in the nasal tract and splitting up of $F1$ into a pole-zero-pole complex (Hawkins and Stevens 1985, cited in

Klatt and Klatt 1990). Klatt and Klatt (1990) also compared these acoustic correlates with the perceptual data. Their results suggest that the relative amplitude of $H1$ and the presence of aspiration noise in the third formant ($F3$) region are relevant to perceiving breathiness. Using synthetic stimuli, Klatt and Klatt concluded that aspiration noise is the most important component of breathy voice. When all cues such as aspiration noise, spectral tilt, OQ, $H1$, and $BW1$ increase together, listeners should perceive increased breathiness and naturalness. For example, many people took $H1$ as the cue of nasality when $H1$ was increased on its own. Therefore, Klatt and Klatt could not conclude the same for natural speech materials, that is, that the aspiration noise is the most reliable cue for perceiving breathiness. Their results for natural speech tokens suggest that the other four cues, such as OQ, spectral tilt, $H1$ ane $BW1$, should be observed together with aspiration noise.

3.6 Summary

The studies summarised in this chapter have contributed to revealing the correlations between physiological and acoustic parameters of breathiness. Owing to the basis of these findings, it seems that measurement of acoustic correlates such as the amplitudes of harmonics and formants, bandwidths of formants, and ratings of noisiness by observation could give us the potential cues for measuring breathiness while avoiding invasive techniques.

However, we need to take into account that there are many difficulties in measuring perceived breathiness, and the various advantages and disadvantages need to be taken into consideration in the analysis of glottal characteristics and perceived breathiness described in the next chapter.

Chapter 4

Politeness: Production and perception (speech of Japanese male speakers)

In this chapter, we report the results of our experimental investigations of the role of voice quality changes in signalling politeness in Japanese, focusing especially on the quantitative analysis of breathiness, and its contribution to the speaker attitude conveyed.

Although this voice quality is typical of female speakers physiologically, it is also observed in males. The communicative role of perceived breathiness has been associated with males' introversion and submissiveness (Moore 1939, Addington 1968, Aronovitch 1976). Furthermore, perceived breathiness is less associated with femininity than the role of high F0 in Japanese (Ohara 1999, 2001). Consequently, it may work as an alternative to the frequency code in expressing deference when a less powerful male speaker addresses a more powerful one.

Moreover, Laver(1980) suggested another probable way in which breathiness may signal politeness, namely by reducing the distance between the speaker and the addressee.

Therefore, breathiness in voice may convey both: 1) deference or submissiveness in males (negative politeness), and 2) closeness between a speaker and a listener (positive politeness). Both can be expressed by a speaker to his/her social superior to achieve negative and positive politeness strategies at a time. In this study, we focus on this "general politeness" function of breathiness, without attempting to decompose it into

the two negative/positive politeness factors.

As noted in Chapter 2, other scholars have suggested some role for paralinguistic cues in Japanese politeness, and there have been a few studies of F0 and speech rate in this connection. A pilot study (Ito 2002) that used tokens extracted from a subset of the dialogues recorded for this study confirmed that 1) speakers do not always change F0 or speech rate according to the relative status of the addressee, and 2) changes of speech rate and F0 do not affect listeners' judgement of the actual relative status of the addressee. We have suggested that these results are not surprising, as there are problems in using either F0 or speech rate to signal politeness: male speakers of Japanese may avoid using high F0 to avoid giving the impression of femininity, and speech rate may convey not only politeness but irritation or boredom. Given these results, and given the reasons for expecting that breathiness may play a role in the paralinguistic signalling of general politeness, the link between breathiness and politeness is our focus here.

In this chapter, I discuss two experiments designed to test the relationship between "general politeness" and breathiness. The first experiment is a production study, involving the Map Task, in which each speaker in dialogue has to cooperate with his senior and junior in turn. The acoustic signal elicited from the spontaneous speech was analysed as a function of speakers' probable intended politeness. The second experiment is a perception test, based on the materials elicited in the first experiment, to determine whether listeners can perceive differences of politeness on the basis of speech alone. Finally, acoustic analyses from the first experiment were correlated with results from the second experiment.

4.1 Speech production test eliciting "general politeness" in natural spontaneous speech

As reviewed in Chapter 2, previous speech perception studies of politeness (e.g. Ogino et al. 1991, Hirose et al. 1997, Ofuka et al. 1999) used stimuli based on production tasks where speakers were explicitly asked to produce utterances in a polite or non-polite way. As a result of the type of production task, the naturalness of the utterances was not guaranteed. This problem needs to be considered seriously to achieve a fair evaluation of expressive speech, as was pointed out by Campbell (2000).

In this study, our goal was to reveal the role of breathiness in expressing politeness in speech produced in a relatively natural manner. Our strategy for collecting natural data was to elicit spontaneous speech in tasks where differences of social status were experimentally varied. In order to elicit natural speech, the Map Task (Anderson et al. 1991, Aono et al. 1994) was used. To extract the effect of general politeness from social status differences only, and to keep social distances between participants as constant as possible, we recruited participants of different social ranks (social superior/inferior pairs) from the same community. This was intended to elicit the adoption of different politeness strategies in dialogue, depending on the social status of the participants. The higher status participant was expected to use either a formal or an informal style, whereas the lower one was expected to use only formal, honorific language (refer to Section 2.3.1).

4.1.1 HCRC Map Task and Japanese Map Task

4.1.1.1 What is the Map Task? (Basic design of the Map Task by HCRC)

The Map Task was originally conceived at the University of Edinburgh, Human Communication Research Centre (HCRC). The Map Task works as follows: the two participants in the dialogue (Instruction giver and Instruction Follower) each have a map with phonologically controlled landmark names. The maps differ slightly in the positions and names of the landmarks. Neither speaker can see the other's map. The Instruction Giver's map has a route marked on it. The task is for the Instruction Giver to explain the direction of the route to the Instruction Follower, referring to the various landmarks along the way.

This must be done accurately enough for the Instruction Follower to draw the route on his or her own map.

Since the Map Task makes participants concentrate on their task, their vocabulary and intentions for every utterance naturally fall within a certain range. This natural limitation of vocabulary and intentions also helps us to collect lexically similar or identical words/phrases without explicit instruction or reading of prepared texts. At the same time the concentration on the task leads the participants to produce natural utterances.

4.1.1.2 Japanese Map Task design for eliciting vocal politeness

For this experiment, all the data were collected using the same maps used for the Japanese Map Task Corpus (Aono et al. 1994, Horiuchi et al. 1997), which is a Japanese version of the HCRC Map Task Corpus (Anderson et al. 1991). For our purposes, the benefits of using the Map Task are as follows. First, the level of formality can be assumed to be constant in all the dialogues between participants. Since the relationship between participants alternates, the effect of relative status change is intended to elicit the production of different voice quality features. Second, the effects of role change (Giver vs. Follower) on voice quality can also be observed. Finally, the Map Task enables us to compare voice quality of lexically similar utterances, for example, extracting frequently used vocabulary, such as "*wakarimasita*" (I understand), "*hidari*" (left), etc. By analysis of these utterances, which occur frequently in the dialogues, it is possible to compare voice quality features.

However, to maximise speakers' use of acoustic features rather than visual markers, eye contact was not allowed between participants. In order to control familiarity, participants including the target speakers were recruited in such a way that each speaker had a conversational partner who belonged to his own working community. This means that familiarity between the participants across dialogues was maintained to a certain extent. However, as mentioned, the relative social status of each participant was controlled so as to elicit general politeness to a social superior. Thus, by examining the dialogues

from speakers when addressing a social superior or inferior conversational partner, we can determine if they use breathiness to show general politeness.

4.1.1.3 Subjects

All five of the subjects were recruited from students and former students of the University of Tokyo. All the subjects were male native speakers of the Tokyo dialect, and their ages ranged from twenty to thirty-four. The subjects were recruited from two different student organisations (Community A and B in Figure 4.1) in the University of Tokyo. However, their background is quite similar, in order to satisfy the conditions of the experiment. Specifically, all speakers have the same gender, dialect and education level, as these are

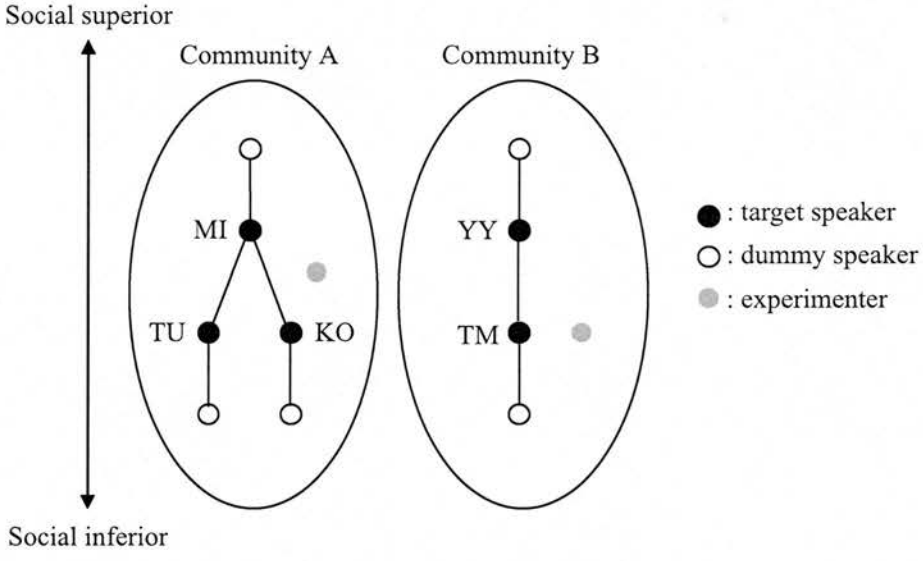


Figure 4.1: Relationship between subjects and an experimenter in the design of production experiment. Note that 1) bold letters indicate IDs of the subjects, 2) a line indicates a pair who held a dialogue, and 3) the experimenter is also a member of their community.

suggested to be fundamental factors of *Keigo* usage by Inoue (1989). The superior/inferior axis in Figure 4.1 reflects the difference of academic years between when they entered the university (e.g. a person who entered in 1988 is superior to a person who entered in 1989). It is the academic year, not the participants' actual age, that controls the relationship in the context of the Japanese university community. To satisfy the condition that the target speakers were able to speak to both their superior and inferior, it was necessary to recruit another five native speakers of the Tokyo dialect. These additional five speakers (dummy,

in Figure 4.1) were not target speakers, but served only as conversational partners in order to satisfy the social status control for pairing of both superior and inferior. To maintain familiarity but to control for status relationship, therefore, a total of ten speakers participated in the recordings. Note that the experimenter is also a member of both communities, that is, the participants were previously familiar with the experimenter. All the participants were told that the experimenter was the only person who would analyse their utterances directly. This helped to minimise the effect of participants speaking more formally because of the presence of an outside observer, as discussed in Chapter 2.

4.1.1.4 Procedures

The speakers were instructed by the experimenter to participate in the task in pairs, made up of a higher status subject and a lower status subject. Each member of the pair would take the role of Instruction Giver or Instruction Follower, and then change roles after completion of one map. They were told that 1) the goal of the task was to draw a Giver's route on the Follower's map, 2) the Giver's and Follower's maps might be different in some respects (*e.g.* Figure 4.2).

The participants performed the task in two separate acoustically insulated rooms where their dialogue was recorded on separate channels. Each participant sat in front of a microphone, wearing a headphone through which they heard their task partner's voice. All materials were digitally recorded (16bit, sampling frequency = 48kHz, stereo) on to DAT tape with a close-talking microphone and a DAT channel for each participant, and were down-sampled to 16kHz for further analysis.

4.1.1.5 Data collection

Dialogues were transcribed in *Hiragana*, a moraic system of writing which is used in Japanese. After being transcribed, selected utterances from the five target speakers were extracted, and were used for further analysis and experiment.

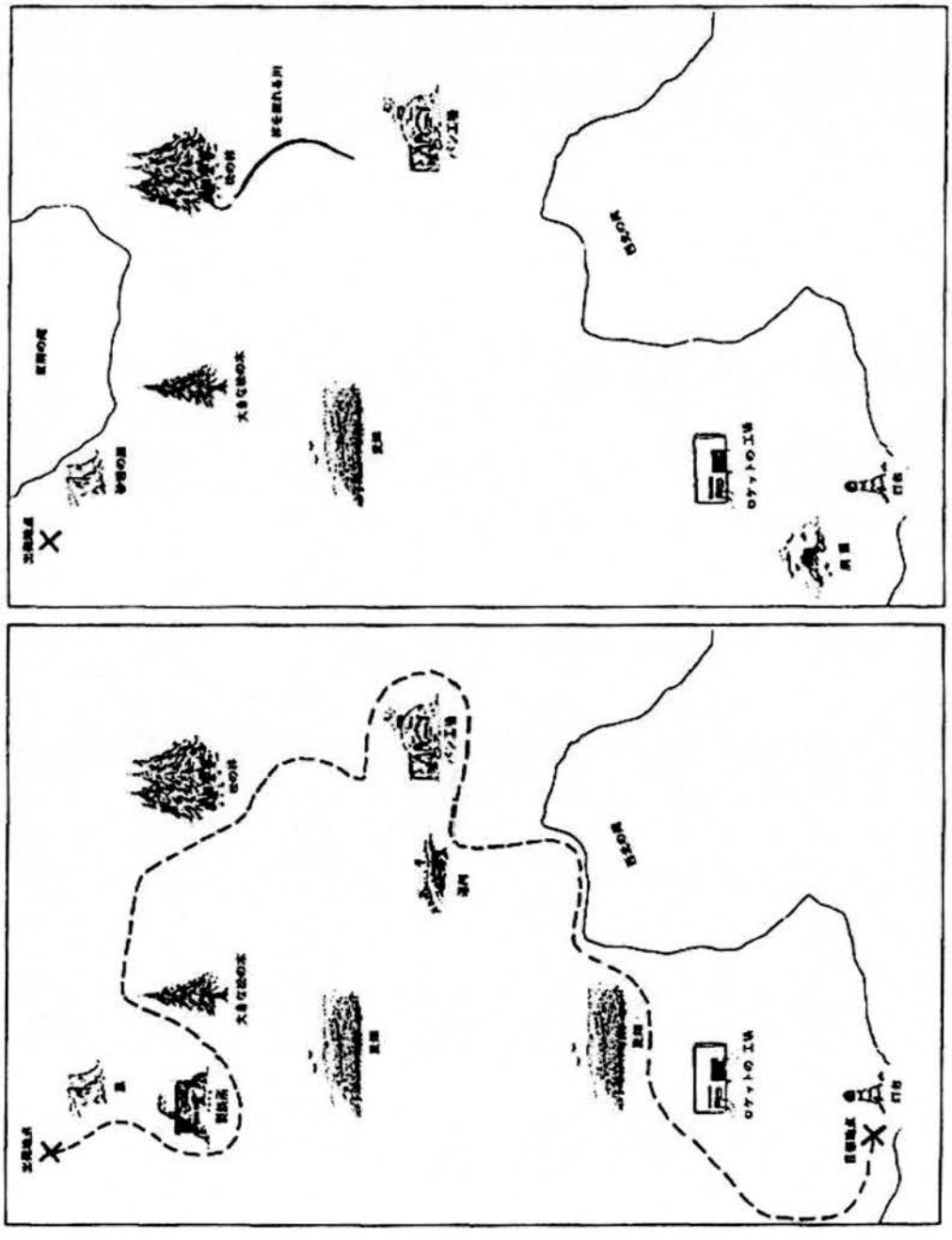


Figure 4.2: Sample maps of Instruction Giver (left) and Instruction Follower (right). The actual size of the maps is A3 (The maps in A3 size are shown in Appendix B).

4.1.2 Analysis of voice quality

In this study, four acoustic parameters of glottal configuration, all of which can be associated with perceived breathiness, were measured directly on the spectra of natural vowels to give indications of vocal fold and glottal configuration. This direct measurement method avoids the disadvantages of invasive methods such as direct physiological measurement (e.g. Rothenberg mask) and of the LF model analysis in which the measures may depend on the experimenter's experience and skills.

As suggested by Hanson et al. (1995, 1997, 1999, 2001) and reviewed in the previous chapter, the direct waveform and spectrum measurements are supported theoretically as follows. First, a change in open quotient (OQ) of the glottal flow waveform affects the behaviour of the spectrum mainly in the lower frequency region. Therefore, the relative amplitude of the first two harmonics ($H1 - H2$) was measured as a reflection of OQ. Second, there may be several sources of change in the spectral tilt of the voicing source (i.e. speed quotient, posterior glottal chinks, and irregular closure of the membranous part of the vocal folds) which result in the decrease in the abruptness with which the airflow through the glottis is cut off, and this causes an increase in spectral tilt. Since increases in spectral tilt are observable mainly in the middle to high frequency region, the relative amplitude of the first harmonic and the third formant frequency ($H1 - A3$) was measured as a reflection of spectral tilt.

Third, the first-formant bandwidth ($BW1$) is affected by the presence and size of the posterior glottal opening. The increase in $BW1$ can be observed in both the speech waveform and the spectrum. In the waveform, the oscillations due to the first formant attenuate more rapidly, and in the spectrum the amplitude of the $F1$ peak is lowered. Therefore, Hanson suggests the following two measures: 1) An estimate of the decay rate of the $F1$ waveform oscillation; 2) The relative amplitude of the first harmonic and the first formant, ($H1 - A1$).

Noise ratings will be described in the next section, because this method involves not only acoustic signal processing but also subjective ratings.

4.1.2.1 Materials for analysis: target tokens

Taking advantage of the potential of the Map Task, frequent expressions which enable us to observe differences in voice quality were marked and extracted.

To facilitate formant analysis, /a/ was selected as the target vowel. Japanese has only five vowels (/a/, /i/, /u/, /e/, /o/) and /a/ has the first and second formants which are most likely to be separable from each other, especially if speakers are males, even with the damping effect of formant peaks in the case of breathy voice.

Syllables showing heavy coarticulation effects from preceding or following consonants (Ní Chasaide, et al. 1993) involving voice quality parameters were avoided . For example, /ka/ was avoided, since /a/ in the syllable can be affected by the breathiness of aspiration of /k/. Also, the nasality of /n/ or /m/ could cause a damping effect on F1 and an increase of $BW1$ on a preceding or following /a/. In the end, the word /hidari/ ('left') was selected. The phonological environment was also controlled so as to have similar F0 contours on all test items. This is due to the fact that identical words have similar pitch patterns, which are lexically specified in Japanese. In the Tokyo dialect, the word /hidari/ has an intonational rise between its first and second mora and it does not contain an accentual fall. Normally, in complete sentences of Japanese, the word /hidari/ would not appear by itself in an utterance final position, because it should be accompanied by a particle or an auxiliary verb; however, in conversation, the word may appear in utterance final position. These utterance final cases were not included in the analysis.

Thirteen to nineteen utterances of the word /hidari/ were collected from each of five speakers. A total of 82 tokens of /hidari/ were extracted from the corpus. These were produced when the speakers had the role of information giver, and when they gave an instruction of direction. This control was supposed to eliminate the possible effects caused by the difference of speech styles which comes from different roles. In addition, the recordings were organised so that the subjects had participated in at least one map task prior to taking the role of information giver. This role and order control was planned to make sure that the participants had enough time to familiarise themselves with the task before taking the information giver's role, which is the target of our recording. The set of

utterances from each speaker contains at least six tokens addressed the speaker's senior (social superior) and at least six tokens addressed to the speaker's junior (social inferior).

4.1.2.2 Acoustic measurements

Acoustic parameter measurements were taken from the middle of the /a/ vowel from "hidari" in cases where /a/ had a duration of at least five pitch periods. Each type of measurement was taken at three positions in each vowel, the duration midpoint of the vowel and 15ms before and after the midpoint. The mean of these three measures per token was computed and used in further analysis.

The parameters measured are summarised for convenience as follows.

BW1 the bandwidth of the first formant

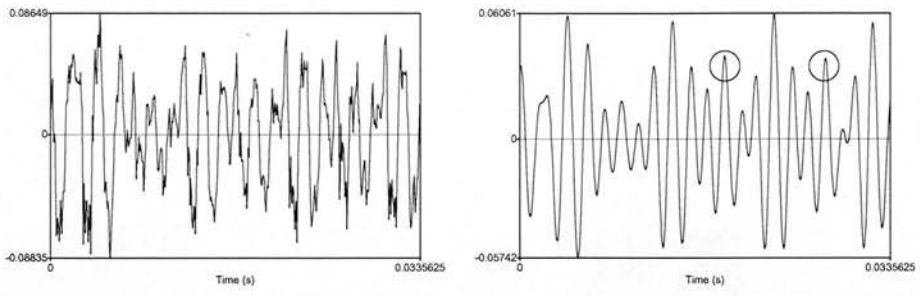
H1-H2 the relative amplitudes of the first and second harmonics, a potential reflection of OQ

H1-A1 the relative amplitudes of the first harmonic and the first formant, a potential reflection of $BW1$

H1-A3 the relative amplitudes of the first harmonic and the third formant, a potential reflection of spectral tilt

There are some differences between these measurement methods and those used by Hanson (1995, 1997, 1999). For the estimation of $BW1$, LPC analysis was employed instead of computing the decay of $F1$ oscillation in the waveform, because there were some tokens which did not have two stable cycles of oscillation, necessary for the estimation of the $F1$ decay. This instability of the oscillation cycle could be a result of the second excitation (Figure 4.3), found mainly in male speakers, because the BW1 estimation by computing decay of oscillations can be easily affected by this second excitation, as pointed out by Hanson (1999).

However, we should also consider the disadvantage of this LPC analysis, which is that it causes smoothing out of the spectrum curve which therefore results in inaccurate estimation of BW1. This is especially true where slopes around the $F1$ peak are more



(a) (b)

Figure 4.3: A possible case of the presence of the 2nd excitation. (a) the original waveform, and (b) the waveform band-pass filtered around $F1$ with a bandwidth of 400Hz

gradual and BW1 is large, BW1 becomes more sensitive to the damping effect of amplitude of $F1$. Nonetheless, LPC analysis is still the most reliable method available.

For estimating formant-frequency and bandwidth with LPC analysis, the Burg algorithm on Praat was used. This algorithm employs a Gaussian-like window instead of a Hamming/Hanning window. The advantage of using a Gaussian-like window is that the Gaussian window is the only window that does not yield frequency peaks due to the shape of the window. For example, if the signal is a sinusoidal wave with a frequency of 1000 Hz, a Hamming window may introduce extra peaks at 950 and 1050 Hz or so, whereas the Gaussian window does not (Boersma, personal communication). Measurement parameters were set as follows: a time step of 10ms and Gaussian-like window length of 50ms, which is equivalent of Hamming window with length of 25 ms (Boersma, personal communication) .

After estimating the formant frequencies and their bandwidths, three other parameters, $H1-H2$, $H1-A1$, and $H1-A3$ were measured and computed. For estimating formant amplitudes, the harmonic peaks nearest the target formant frequencies in the FFT spectra were measured manually. The size of the FFT is determined as follows. In the case of a Gaussian-like window, the window length is the equivalent of a Hamming window multiplied by 2 (Boersma, personal communication). Thus, if we would like to compare with the results of Hamming window with length of 16 ms, the window length becomes

32 ms, which has 500 samples if the sampling frequency is 16kHz. Then the nearest power of two to 500 was chosen as the number of sampling points (Boersma, personal communication). The FFT in this case has been done for 512 ($= 2^9$) sampling points. In this study, the following parameters were measured (Figure 4.4): 1) the amplitude of the first harmonic ($H1$); 2) the amplitude of the second harmonic ($H2$); 3) the harmonic nearest from the first formant ($A1$); and 4) the harmonic nearest from the third formant ($A3$).

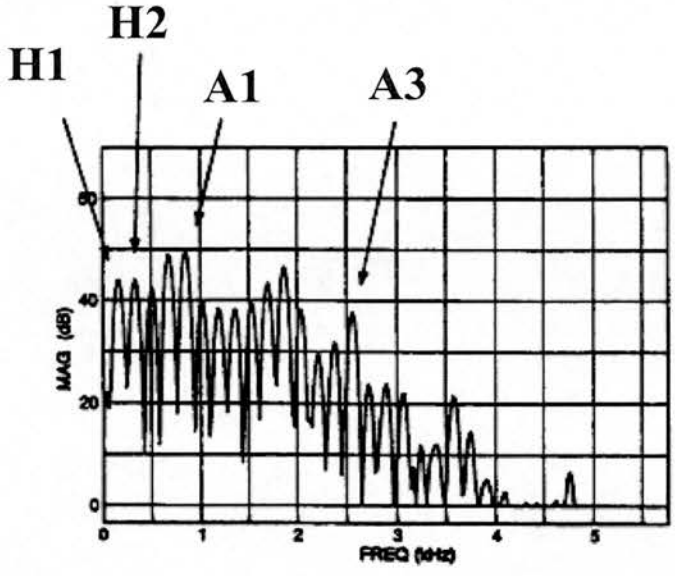


Figure 4.4: The measurement of acoustic parameters in FFT Spectra (From Hanson 1999)

To allow comparison across speakers, the modification suggested by Hanson (1995, 1997) was applied for the measured $H1$, $H2$, and $A3$. The quantity,

$$20 \log_{10} \left[F1^2 / \left(F1^2 - f^2 \right) \right]$$

is subtracted from $H1$ and $H2$, where f is the frequency at which the harmonic is located. The quantity

$$20 \log_{10} \left(\frac{\left[1 - \left(\frac{F3}{F1} \right)^2 \right] \left[1 - \left(\frac{F3}{F2} \right)^2 \right]}{\left[1 - \left(\frac{F3}{\widetilde{F1}} \right)^2 \right] \left[1 - \left(\frac{F3}{\widetilde{F2}} \right)^2 \right]} \right)$$

is added to A3, where $\widetilde{F1}$ and $\widetilde{F2}$ are the first and second formant frequencies of a neutral vowel, estimated from the mean of the third formant frequencies.

This modification converts the measured H1, H2, and A3 into $H1^*$, $H2^*$, and $A3^*$, which are equivalent values of the average speaker's neutral vowel. Thus, $H1^*$, $H2^*$, and $A3^*$ were obtained by this modification.

Hanson (1995) also suggested an additional correction to compensate for the amplitude of the third formant due to the difference between the bandwidth across vowels which was reported by House and Stevens (1958) (e.g. the A3 value of /æ/ is expected to be 1dB smaller than that of /ɛ/ for male speakers). However, this correction was not employed in this study, because the vowel /a/ is the only target vowel used for the stimuli and it only gives the same correction across stimuli.

4.1.2.3 Noise ratings from the waveform

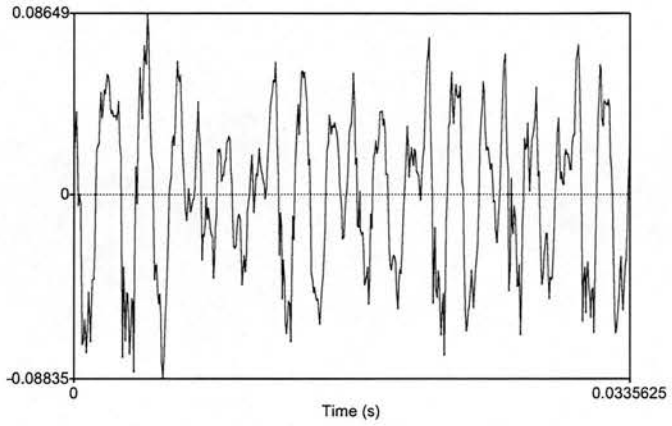
In addition to the parameters measured in the previous section, a noise rating test was introduced. Three students studying linguistics or music who had taken an acoustics class for more than two terms participated as raters in this test. In this method, the vowels were bandpass-filtered around the third formant using a Hann band filter that has the shape of a raised sine (a Hann window) in the frequency domain, with a bandwidth of 400 Hz. This bandwidth is narrower than that chosen by Klatt et al. (1990) and Hanson (1995, 1997), because the mean of $F3$ across speakers is 2390Hz, which is much lower than in Hanson's data (e.g. 555 * 5 = 2775 (Hz), estimated from Hanson 1995, 1997). To filter out the interference of neighbouring formants, this narrower bandwidth was chosen.

The bandpass-filtered waveforms corresponding to the speech segments (the /a/ vowel from the word /hidari/ obtained from the Map Task recordings) used in the previous spectral measurement were given as graphic presentation to the raters. The waveform diagrams were prepared by the experimenter so as to contain at least five to seven pitch periods, and were adjusted so that their amplitudes had been scaled to fill the full available range. They were then presented to the raters who were asked to rate the stimuli according to the irregularity of the waveforms, that is, the amount of noise on a scale from "0" to "10", where"0" means essentially no evidence of noise interference and

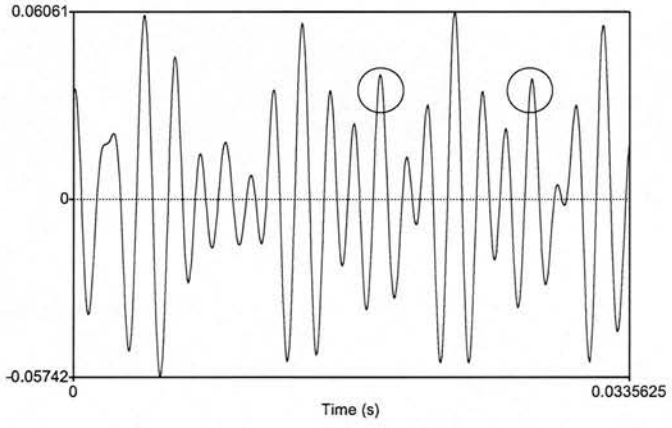
"10" means little evidence of periodicity.

These judgements were made independently by each rater. Their ratings were highly correlated ($r = 0.94$ for raters 1 and 2, $r = 0.83$ for raters 1 and 3, and $r = 0.72$ for raters 2 and 3). For each stimulus, the mean of ratings from the raters was computed as the noise rating value (Nw).

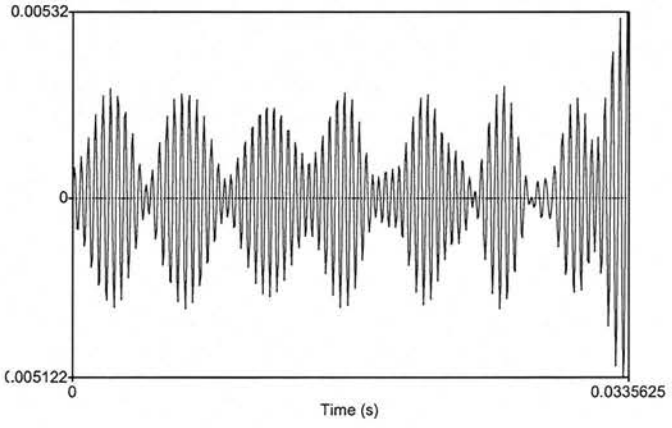
The concern about the noise rating method, as stated by Hanson et al. (1999), is that the ratings may be influenced by the presence of the second excitation (Figure 4.5), which is more likely to be found in male speakers' voices. Therefore, the presence of the second excitation for each stimulus was also observed.



(a) the original waveform



(b) bandpass-filtered around F1



(c) bandpass-filtered around F3

Figure 4.5: A possible case of the presence of the 2nd excitation. (a) the original waveform, (b) the waveform bandpass-filtered around F1 with a bandwidth of 400Hz, (c) the waveform bandpass-filtered around F3 with a bandwidth of 400Hz

Perturbation quotient in the high frequency region In the noise rating method, if we extract the component of a waveform bandpass-filtered around the $F3$ region and make visual regularity ratings of the waveforms, there might be a concern that these ratings might give subjective values, rather than quantitative measures. Therefore, in addition to the measures suggested by Hanson (1995, 1997), this study employs another quantitative method of measuring turbulent noise. If we compare the results of noise ratings and that of other acoustic measurements of waveform irregularity, using the extracted tokens, we can check whether the results of the measurement quantitatively reflect the irregularity of the waveform from aspiration noise by computing the correlation between the quantitative measures of turbulent noise and Nw.

The Harmonic-to-Noise ratio (HNR) was found to be associated with Nw (Hanson 1995) but this measure could be determined only by estimating the glottal source and aspiration with synthesised stimuli. In natural spontaneous speech, it is quite difficult to separate the noise from the periodic component and to measure each separately (Hanson 1995). HNR is also influenced by F0 perturbations, such as jitter (for the details, refer to Qi et al. 1997, Murphy 1999, 2000).

To examine the aspiration noise, a method for measuring F0 and amplitude perturbation of the waveform from the high frequency region was employed.
The arguments among voice quality scientists are as follows. Hanson (1995) mentioned that measures such as jitter (changes in F0) and shimmer (changes in amplitude of excitation) reflect the irregularity of vocal fold vibration, whereas aspiration noise theoretically arises from a posterior chink in the glottis. On the other hand, many researchers (e.g. Takahashi and Koike 1975, cited in Buder 2000; Shrivastav 2001, 2003) found that perceived breathiness is significantly correlated with objective measures such as F0 and amplitude perturbations. Although it should be taken into account that some of their studies were based on pathological voices, the perturbation measures which represent the irregularity of the waveform were found to have strong and positive associations with perceived breathiness. If we consider that our interest is aspiration noise in the high frequency region, it seems reasonable to measure these perturbation

quotients from the band-pass filtered waveform around the $F3$ region, to find out if F0 and/or amplitude perturbation in the high frequency region could be acoustic cues for aspiration noise and breathiness.

For the measurement of F0 and amplitude perturbation, Praat (version 4.1.9) functions were employed for all the tokens band-pass filtered around the $F3$ region with the bandwidth of 400Hz (the same condition as used for noise ratings). F0 perturbation was measured as follows. The Relative Average Perturbation (RAP), which is the average absolute difference between a pitch period and the average of it and its two neighbour periods, was divided by the average period.

Amplitude perturbation was measured as follows. The three-point Amplitude Perturbation Quotient (APQ3), which is the average absolute difference between the amplitude of a pitch period and the average of the amplitudes of its neighbour periods, was divided by the average amplitude. For pathological studies, APQ11 (the eleven-point APQ) is likely to be used for measurement of shimmer. However, the purpose of this study is to observe natural spontaneous speech, and obtaining eleven stable cycles of the target vowel is extremely difficult because of the lack of sustained vowels. Therefore, APQ3 was employed as an available option.

Additionally, RAP and APQ3 of the original waveform were measured. By comparing the RAP and APQ3 values of the original waveform and of the waveform after it was bandpass-filtered around the $F3$ region, we are able to observe how aspiration noise in the high frequency region may contribute to the measurement of the irregularity of waveforms.

4.1.2.4 Summary of spectral parameters and noise parameters

Here is a summary of acoustic parameters employed for the analysis of production test. The parameters are divided into two groups, one focusing on spectral characteristics (spectral-related parameters) and the other based on aspiration noise (noise-related parameters).

The spectral-related parameters are as follows.

***BW1*:** the bandwidth of the first formant, measured by LPC analysis

H1*-H2*; the relative amplitudes of the first and second harmonics, a potential reflection of OQ

***H1*-A1*:** the relative amplitudes of the first harmonic and the first formant, a potential reflection of BW1

H1*-A3*: the relative amplitudes of the first harmonic and the third formant, a potential reflection of spectral tilt

The following parameters are noise-related parameters.

***Nw*:** the ratings of irregularity of the waveform after bandpass-filtered in the third formant region

***RAP_F3*:** the F0 perturbation quotient of the waveform after bandpass-filtered in the third formant region

***APQ3_F3*:** the amplitude quotient of the waveform after bandpass-filtered in the third formant region

4.1.3 The aim of acoustic analysis of production

This part of the study was intended to answer the following questions.

1. Do speakers change their voice quality according to the relative status of the addressee to express their politeness?
2. If so, do they try to change their voice in a way to increase the degree of perceived breathiness so as to express "general politeness" (simultaneous deference and closeness) to their social superior?
3. Do all speakers change their voice in a similar way or not?

To answer these questions, the following statistic analyses were conducted, employing acoustic and noise parameters (refer to the previous section).

First of all, to observe the overall tendency, the mean and standard deviation of the acoustic parameters obtained from the stimuli addressed the social superior and the social inferior were computed. To check whether the means and standard deviations of the stimuli addressed superior and the stimuli addressed inferior are significantly different or not, analyses of variance (ANOVA) were performed per speaker. If all the speakers changed the voice in a similar way, then the mean and the significant level ANOVAs would show the same tendency across the speakers, depending on the stimulus groups of the relative status of the addressee. On the other hand, if the speakers changed their voices in different ways, we need to examine each case in order to see what kind of changes each speaker made.

Pearson product moment correlation coefficients between the acoustic parameters across the speakers were computed so as to observe if the acoustic parameters relevant to the glottal characteristics such as the size of a posterior chink would correlate with each other, as suggested by Hanson (1995). Also this will give us the information how the correlation between the acoustic parameters may change according to the relative status of the addressee, and to what extent the correlation between the acoustic parameters depends on speaker characteristics.

4.1.4 Results

This section reports the acoustic characteristics of individual speakers, obtained from the production task. The acoustic parameters (refer to section 4.2.2.4) will be discussed particularly in relation to the relative status of each addressee.

Firstly, the mean and standard deviation of the acoustic parameters of speakers are computed, to observe speaker-dependent characteristics of these parameters, regarding the status of the addressee of each token. Table 4.2 and Figure 4.6 show the mean and standard deviation of acoustic parameters by speakers, associated with the relative status of the addressee. Analyses of variance (ANOVA) are computed as well (Table 4.1).

Table 4.1 shows that there are significant effects of the relative status of the addressee, but that the effects are speaker-specific. For example, the result of ANOVA shows a difference in $H1^* - H2^*$ between the addressees of different status for Speaker MI. For Speaker KO, the results of ANOVAs show the difference in $BW1$, Nw, RAP_F3, and APQ_F3 between the addressees of different status. For Speaker TM, the result of ANOVAs shows the difference in Nw between the addressees of different status. For Speaker YY, the result of ANOVAs shows the difference in $H1^* - A1$ between the addressees of different status. As for Speaker TU, there was no significant difference observed.

Table 4.1: Results of analyses of variance (ANOVAs) performed to examine differences in acoustic parameters across status of addressee per speakers. (Bold face indicates significance ($p < 0.05$))

Speaker	Df	BW1	H1*-H2*	H1*-A1	H1*-A3*	Nw	RAP_F3	APQ3_F3
MI	F(1,12)	.006	**4.967**	1.757	.380	.046	.379	.112
	(Sig.)	(.940)	**(.048)**	(.212)	(.550)	(.834)	(.551)	(.744)
TU	F(1,15)	.123	.240	.006	2.135	.073	.071	.854
	(Sig.)	(.731)	(.632)	(.939)	(.166)	(.790)	(.794)	(.371)
KO	F(1,17)	**4.675**	.033	.132	.030	**4.133**	**23.14**	**9.944**
	(Sig.)	**(.046)**	(.858)	(.721)	(.864)	**(.049)**	**(.000)**	**(.006)**
TM	F(1,18)	.223	.151	.806	1.545	**6.398**	1.309	1.334
	(Sig.)	(.643)	(.703)	(.382)	(.231)	**(.022)**	(.268)	(.264)
YY	F(1,15)	2.453	1.363	**6.190**	.000	.660	.189	.005
	(Sig.)	(.140)	(.263)	**(.026)**	(.996)	(.430)	(.671)	(.945)

4.1.4.1 Correlations between spectral parameters ($H1^* - H2^*$, $H1^* - A1$, $H1^* - A3^*$)

The correlation between the acoustic parameters was checked across speakers so as to observe how the acoustic parameters depend on speakers' individual physiological characteristics and/or the relative status of the addressee.

The mean and standard deviation of spectral parameters per speaker according to the relative status of the addressee are computed (Tables 4.1, 4.2, Figure 4.6). Scatter-plot diagrams show correlation between spectral parameters (Figures 4.7- 4.9). Tables showing the results of Pearson product moment correlation coefficients are also given (Tables 4.3, 4.4). The results of spectral measurement show that speakers' physiological

Table 4.2: Mean and standard deviation of acoustic parameters per speaker according to addressee status, where L-H indicates that the addressee is superior and H-L indicates that the addressee is inferior, Nw is the waveform-based noise judgements, and RAP_F3, and APQ3_F3 were bandpass-filtered centered at $F3$ with a bandwidth of 400Hz. (Bold face indicates significance ($p < 0.05$), by ANOVA test (Table 4.1))

Speaker	Status (No. of tokens)	$BW1$ (Hz) Mean (s.d.)	H1*-H2* (dB) Mean (s.d.)	H1*-A1 (dB) Mean (s.d.)	H1*-A3* (dB) Mean (s.d.)	Nw Mean (s.d.)	RAP_F3 (%) Mean (s.d.)	APQ3_F3 (%) Mean (s.d.)
MI	H-L (N=6)	69.9 (10.7)	**7.74 (2.08)**	-0.20 (2.19)	29.9 (2.64)	6,33 (1.62)	0.226 (0.063)	2.52 (0.66)
	L-H (N=7)	70.7 (19.9)	**5.36 (1.78)**	-2.16 (2.98)	28.9 (3.26)	6.00 (1.54)	0.281 (0.209)	2.82 (2.07)
TU	H-L (N=8)	59.9 (45.3)	3.34 (4.89)	-8.57 (5.55)	25.4 (10.03)	7.13 (1.66)	0.396 (0.381)	3.34 (1.26)
	L-H (N=8)	53.6 (21.8)	4.38 (3.48)	-8.41 (1.74)	18.8 (7.78)	7.13 (1.31)	0.445 (0.351)	4.09 (1.90)
KO	H-L (N=9)	**34.1 (14.7)**	4.79 (4.17)	-6.28 (1.77)	25.3 (6.92)	**7.15 (1.04)**	**0.937 (0.327)**	**6.52 (3.81)**
	L-H (N=9)	**55.9 (26.3)**	4.49 (2.57)	-5.68 (4.67)	24.7 (8.03)	**5.70 (1.86)**	**0.334 (0.185)**	**2.31 (1.26)**
TM	H-L (N=9)	98.4 (29.5)	0.31 (1.84)	-7.67 (2.10)	10.0 (4.55)	**6.41 (1.91)**	0.400 (0.117)	5.45 (2.63)
	L-H (N=10)	104.8 (29.6)	0.63 (1.69)	-6.49 (3.39)	13.2 (6.30)	**8.23 (1.20)**	1.380 (0.256)	4.31 (1.62)
YY	H-L (N=8)	73.5 (25.7)	-2.07 (2.21)	**-14.07 (1.83)**	7.53 (4.30)	6.75 (1.68)	0.376 (0.502)	4.75 (3.34)
	L-H (N=8)	55.5 (19.9)	-1.11 (0.74)	**-16.78 (2.49)**	7.54 (4.01)	5.54 (1.83)	0.485 (0.500)	4.65 (2.26)

characteristics seem to be a dominant factor associated with perceived breathiness, due to the fact that strong correlation between spectral parameters were observed (Table 4.3) and the range of spectral parameters was found to be speaker-specific (Figures 4.6, 4.8, 4.9).

Speaker MI shows large $H1^* - H2^*$ (a possible indicator of OQ) and $H1^* - A3^*$ (a possible indicator of spectral tilt), possibly resulting from a larger OQ, and steeper spectral tilt, possibly resulting from a larger glottal chink and constant air leakage at the glottis. By contrast, Speaker YY shows the opposite tendency to Speaker MI in the spectral characteristics, which might suggest having a smaller glottal chink and almost no air leakage at the glottis. These two speakers do not have large overlaps in the ranges of their spectral parameters so these two speakers having opposite tendencies in spectral

parameters can be categorised as two distinct cases of opposite glottal characteristics. The other three speakers (Speakers TU, KO, TM) do not show clear tendencies which assist in determining their glottal characteristics. It is difficult to estimate these three speakers' glottal condition. $H1^* - A3^*$ is strongly and positively correlated with Nw for Speaker TM, who has larger $BW1$ values.

We might possibly say that some of the spectral parameters obtained from five speakers show strong and positive associations with each other across speakers (Table 4.3). From the observation of scatter-plot diagrams of spectral characteristics which seem to be relatively clustered by speakers (Figures 4.7- 4.9), the spectral characteristics related to glottal characteristics might be speaker-dependent, and yet follow approximately the same positive correlation between spectral parameters. Another general tendency found (with the exception of Speaker TU) is that when participants speak to their social inferior, the correlation between $H1^* - A1$ and $BW1$ is weak, whereas when they speak to their social superior, the correlation is positive and strong (Table 4.4). This strong positive correlation between $H1^* - A1$ and $BW1$ observed in the tokens addressed to the speakers' social superiors implicates the contribution of breathiness in speech addressed to their social superiors, considering the fact that when the degree of breathiness changes all the spectral parameters change together accordingly.

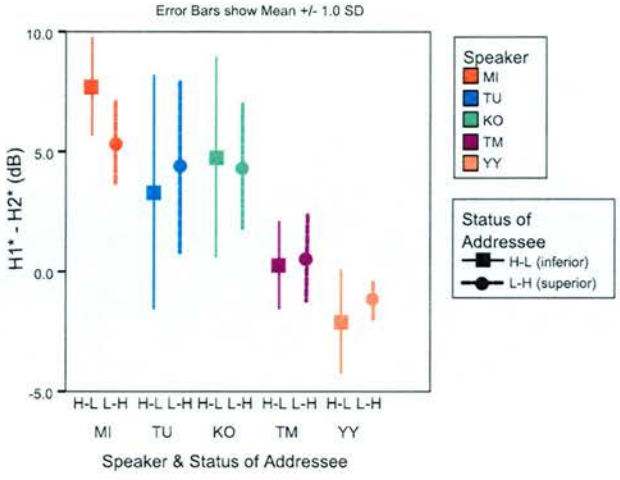
As for the relative status of addressee, no significant difference was found between the acoustic measures from tokens addressed the social superior and those from tokens addressed the social inferior, except $H1^* - H2^*$ of Speaker MI and $H1^* - A1$ of Speaker YY. From this result, it is difficult to say that speakers changed their glottal characteristics according to the status of the addressee. Even if the change happens, it is manifested in different ways for different speakers. However, our data suggest that each speaker is largely constrained by the controllability range of voice change due to physical properties of their articulatory organs, and we therefore might expect any effects of addressee status to be subtle at best. We therefore argue that these small changes in

spectral characteristics are indicative of genuine effects.

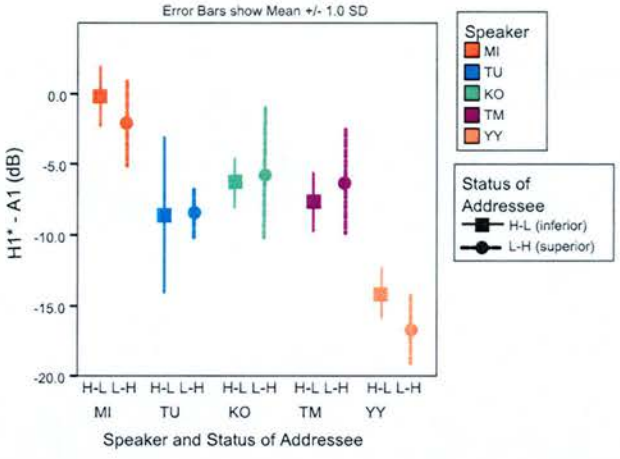
Table 4.3: Pearson product moment correlation coefficients and significance level between acoustic parameters, where tables are provided for all the speakers. (Correlation is significant at *: 0.05, **: 0.01 level (2-tailed).)

Acoustic Parameters	BW1 (Hz)	H1*-H2* (dB)	H1*-A1 (dB)	H1*-A3* (dB)	Nw	RAP_F3 (%)	APQ_F3 (%)
BW1(Hz)	1	-.027	.296(**)	-.279(*)	.145	-.078	-.055
H1*-H2*(dB)		1	.675(**)	.569(**)	.226(*)	-.108	-.276(*)
H1*-A1(dB)			1	.623(**)	.411(**)	.011	-.138
H1*-A3*(dB)				1	.212	-.006	-.155
Nw					1	.105	.201
RAP_F3(%)						1	.390(**)
APQ3_F3(%)							1

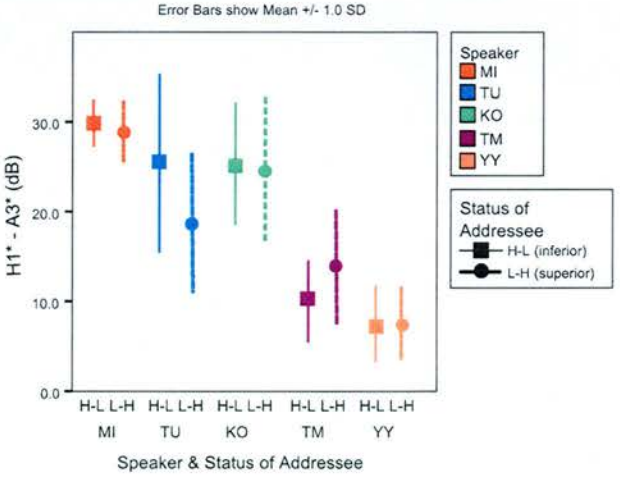
Across speakers, some general tendencies are found. $H1^* - H2^*$ is positively correlated only with $H1^* - A1$ and $H1^* - A3^*$, at the same time, there is a relatively strong positive correlation between $H1^* - A1$ and $H1^* - A3^*$. Therefore these three spectral parameters seem to be dominated by $H1^*$. $H1^* - A1$ is weakly but significantly correlated with $BW1$, which partly supports the prediction that $H1^* - A1$ is a possible indicator of $BW1$ (Hanson, 1995). However the correlation is not strong. Considering the results from Tables 4.2 - 4.4, it seems that $H1^* - A1$ and $BW1$ are positively correlated with each other when the speakers addressed the social superior (Figure 4.7). However, as we can observe from the scatterplot, it is worth noting that the distribution of $BW1$ seems to have wide ranges with overlaps between the speakers, whereas the distribution of $H1^* - A1$ seems to be clustered by each speaker.



(a)



(b)



(c)

Figure 4.6: Means (points) and standard deviations (lines) to compare (a) $H1^* - H2^*$, (b) $H1^* - A1$, and (c) $H1^* - A3^*$ across speakers and the relative status of the addressee. (H-L: the addressee is the social inferior, L-H: The addressee is the social superior)

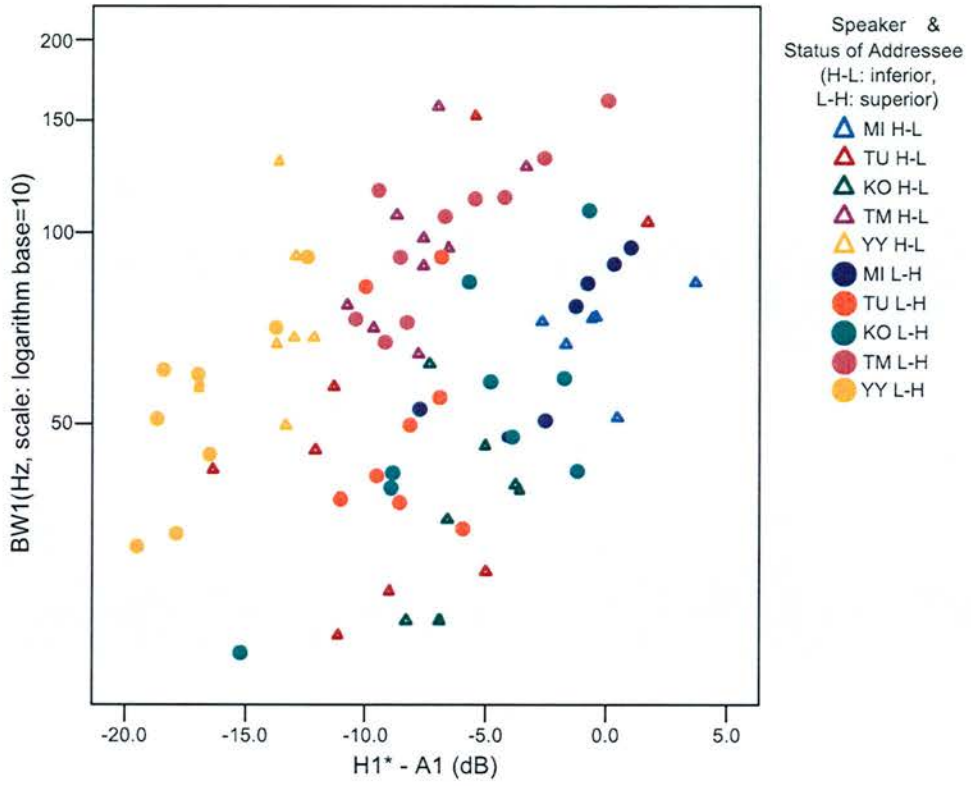


Figure 4.7: Relation between $BW1$ (logarithmic scale) and $H1^* - A1$. Data points for the cases when the addressee is inferior to the speaker are displayed as open triangles.

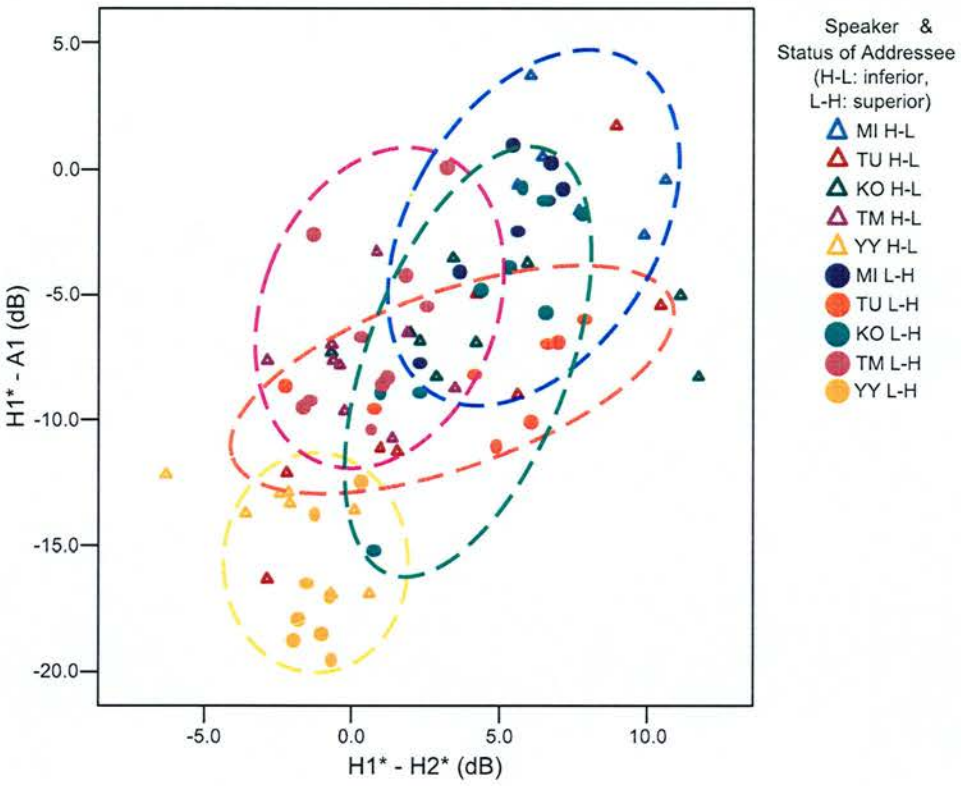


Figure 4.8: Relation between $H1^* - H2^*$ and $H1^* - A1$. Data points for the cases when the addressee is inferior to the speaker are displayed as open triangles. Dashed lines show areas of case clustering per speaker.

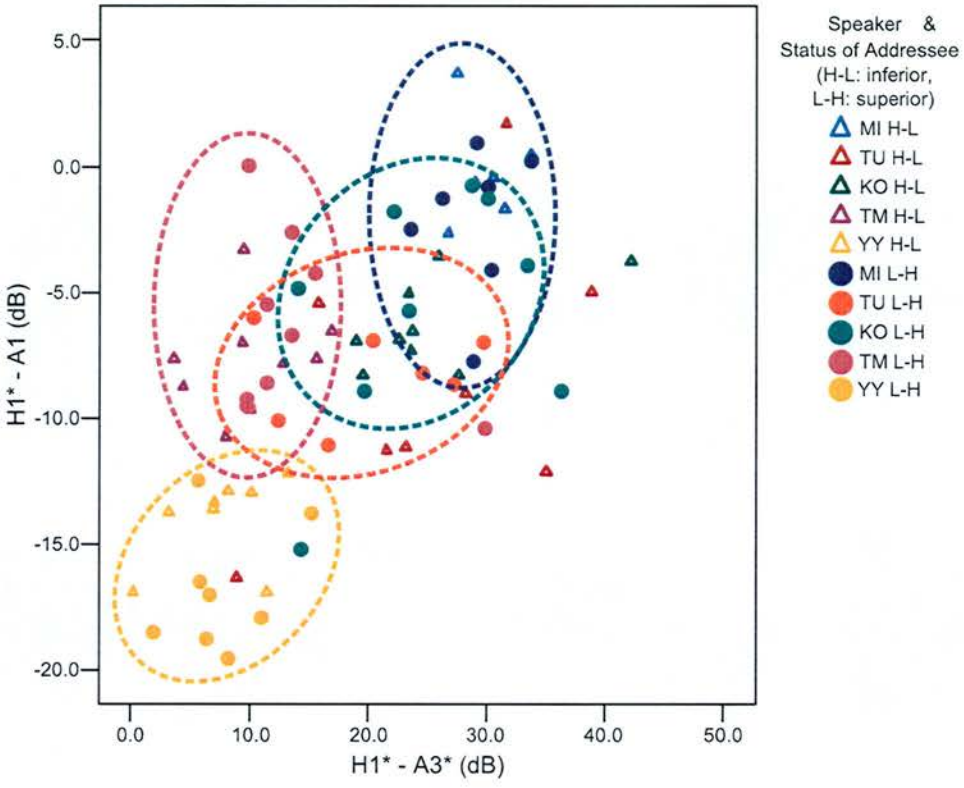


Figure 4.9: Relation between $H1^* - A1$ and $H1^* - A3^*$. Data points for the cases when the addressee is inferior to the speaker are displayed as open triangles. Dotted lines show areas of case clustering per speaker.

Table 4.4: Pearson product moment correlation coefficients between acoustic parameters with significance level, where tables are provided per each speaker(sp). (Correlation is significant at *: 0.05, **: 0.01 level (2-tailed).)

sp.	status	parameter	*BW*1	H1*-H2*	H1*-A1	H1*-A3*	*Nw*	RAP_F3	APQ3_F3
MI	H-L	*BW*1	1	.097	.312	-.842(*)	.956(**)	.036	-.732
	(N=6)	H1*-H2*		1	-.538	-.149	.375	.029	-.425
		H1*-A1			1	-.052	.180	-.193	-.166
		H1*-A3*				1	-.788	.214	.667
		Nw					1	.040	-.792
		RAP_F3						1	-.293
		APQ3_F3							1
	L-H	*BW*1	1	.683	.826(*)	.432	.615	-.401	-.371
	(N=7)	H1*-H2*		1	.863(*)	.042	.535	-.393	-.206
		H1*-A1			1	.189	.659	-.203	-.042
		H1*-A3*				1	.551	.353	.235
		Nw					1	.360	.392
		RAP_F3						1	.962(**)
		APQ3_F3							1
TU	H-L	*BW*1	1	.700	.532	-.276	.313	-.047	.195
	(N=8)	H1*-H2*		1	.850(**)	.134	.484	.249	.580
		H1*-A1			1	.491	.267	.331	.596
		H1*-A3*				1	-.289	.524	.474
		Nw					1	.044	.163
		RAP_F3						1	.902(**)
		APQ3_F3							1
	L-H	*BW*1	1	.429	.089	.032	-.404	-.305	-.333
	(N=8)	H1*-H2*		1	.396	-.201	.318	-.883(**)	-.850(**)
		H1*-A1			1	.250	.039	-.193	-.556
		H1*-A3*				1	-.065	.060	.124
		Nw					1	-.443	.036
		RAP_F3						1	.697
		APQ3_F3							1
KO	H-L	*BW*1	1	-.367	.425	.271	-.099	-.014	-.553
	(N=9)	H1*-H2*		1	.094	-.061	-.445	-.095	-.232
		H1*-A1			1	.580	.453	.635	-.389
		H1*-A3*				1	.178	.753(*)	.021
		Nw					1	.518	.382
		RAP_F3						1	.157
		APQ3_F3							1
	L-H	*BW*1	1	.586	.636	.138	.021	.117	.210
	(N=9)	H1*-H2*		1	.871(**)	.273	.383	.313	.482
		H1*-A1			1	.394	.391	.174	.347
		H1*-A3*				1	.246	.068	-.099
		Nw					1	.260	.405
		RAP_F3						1	.834(**)
		APQ3_F3							1
TM	H-L	*BW*1	1	.003	.563	-.213	-.126	.087	-.027
	(N=9)	H1*-H2*		1	-.077	.062	-.123	-.219	-.398
		H1*-A1			1	.218	.592	.211	.164
		H1*-A3*				1	.644	.657	-.199
		Nw					1	.552	.266
		RAP_F3						1	.119
		APQ3_F3							1
	L-H	*BW*1	1	.327	.866(**)	-.243	.140	-.371	-.042
	(N=10)	H1*-H2*		1	.476	.015	.097	.101	-.286
		H1*-A1			1	-.254	.161	-.160	.143
		H1*-A3*				1	.598	-.319	-.140
		Nw					1	-.701(*)	.423
		RAP_F3						1	-.335
		APQ3_F3							1
YY	H-L	*BW*1	1	.223	.311	.039	-.361	-.137	.180
	(N=8)	H1*-H2*		1	-.699	-.516	-.593	-.723(*)	-.563
		H1*-A1			1	.408	.270	.326	.359
		H1*-A3*				1	-.028	.600	.535
		Nw					1	.714(*)	.629
		RAP_F3						1	.930(**)
		APQ3_F3							1
	L-H	*BW*1	1	.631	.830(*)	-.094	.722(*)	.355	-.001
	(N=8)	H1*-H2*		1	.519	-.234	.353	.456	-.031
		H1*-A1			1	.299	.731(*)	.146	-.099
		H1*-A3*				1	-.205	-.314	.012
		Nw					1	.521	.290
		RAP_F3						1	.691
		APQ3_F3							1

4.1.4.2 Correlations between aspiration noise related parameters

Here, the results of potential indicators of aspiration noise observed in the third formant frequency region are reported. Our interest is the aspiration noise in the high frequency region, which could be an attribute of perceived breathiness. Results are shown in the tables (Tables 4.2,4.5) and figure (Figure 4.10) which represent the computed mean and standard deviation of noise-related parameters, scatter-plot diagram between these noise-related parameters (Figure 4.11), and correlation tables of Pearson product moment correlation coefficients (Tables 4.3- 4.4).

For Speaker KO, there are strong associations between the relative status of the addressee and F0 perturbation (RAP_F3) and amplitude quotient (APQ_F3), such that he has larger RAP_F3 and APQ_F3 when he speaks to his social inferior. Also this speaker shows a moderate association between the direction of addressee and Nw. A striking fact is that Speaker KO shows the opposite tendency to that of the other speakers, in associating the relative status of his addressee and F0 and amplitude perturbations in the $F3$ region. Overall, the other speakers have larger RAP_F3 and APQ_F3 when they addressed their social superiors.

To check whether the irregularity of F0 and amplitude of Speaker KO came from aspiration noise or aperiodic vocal fold vibration, the F0 and amplitude perturbation quotients of the original waveform are consulted. Table 4.5 shows the mean and standard deviation of RAP and APQ3 obtained from the original waveform and the waveform bandpass-filtered around $F3$, and the differences between these two quotients and the filtered waveform (RAP-RAP_F3, APQ3-APQ3_F3, respectively), according to the relative status of the addressee, marked with significance level of one-way ANOVA. From this result, Speaker KO shows irregularity of F0 and amplitude overall in the original waveform as well as in the waveform bandpass-filtered around the $F3$ region. As was mentioned above, he does not seem to have used aspiration noise according to the addressee status. The association with Nw found in this speaker could be suspected from this irregularity as well, because noise rating method is also influenced by any

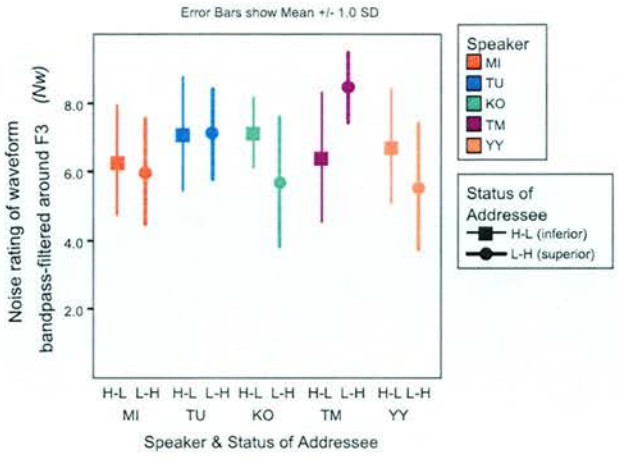
Table 4.5: The means and standard deviations (the values in parentheses) of F0 and amplitude perturbation of original waveform (RAP and APQ3, respectively), F0 and amplitude perturbation bandpass-filtered around F3 (RAP_F3, APQ_F3, respectively), and the difference of these two quotients, according to the relative status of addressee (bold face indicates significance level ($p < 0.01$, by ANOVA test.)

speaker (token)	acoustic param. direction	RAP Mean (s.d.)	APQ3 Mean (s.d.)	RAP_F3 Mean (s.d.)	APQ3_F3 Mean (s.d.)	RAP-RAP_F3 Mean(s.d.)	APQ3-APQ3_F3 Mean(s.d.)
MI	H-L	0.274 (0.096)	1.901 (0.735)	0.226 (0.063)	2.521 (0.663)	-0.975 (0.200)	-0.141 (0.259)
(N=13)	L-H	0.285 (0.140)	1.477 (0.751)	0.281 (0.751)	2.817 (2.069)	-0.953 (0.191)	-0.263 (0.275)
TU	H-L	0.371 (0.389)	3.445 (1.670)	0.396 (0.381)	3.343 (1.256)	-1.098 (0.302)	-0.009 (0.098)
(N=16)	L-H	0.404 (0.260)	3.108 (1.957)	0.445 (0.351)	4.087 (1.899)	-1.03 (0.235)	-0.126 (0.220)
KO	H-L	**0.752 (0.282)**	**5.479 (2.890)**	**0.937 (0.327)**	**6.523 (3.805)**	-0.914 (0.197)	-0.071 (0.169)
(N=18)	L-H	**0.294 (0.177)**	**2.571 (1.390)**	**0.334 (0.185)**	**2.311 (1.256)**	-0.877 (0.246)	0.087 (0.223)
TM	H-L	0.455 (0.302)	2.951 (1.792)	0.400 (0.117)	5.451 (2.628)	-1.096 (0.242)	-0.287 (0.301)
(N=19)	L-H	0.535 (0.404)	2.752 (1.001)	1.380 (2.560)	4.309 (1.617)	-0.967 (0.329)	-0.197 (0.234)
YY	H-L	0.571 (0.652)	4.675 (3.023)	0.376 (0.502)	4.746 (3.344)	-1.078 (0.297)	0.001 (0.219)
(N=16)	L-H	0.392 (0.478)	5.704 (3.993)	0.485 (0.500)	4.646 (2.251)	-1.182 (0.264)	0.046 (0.256)

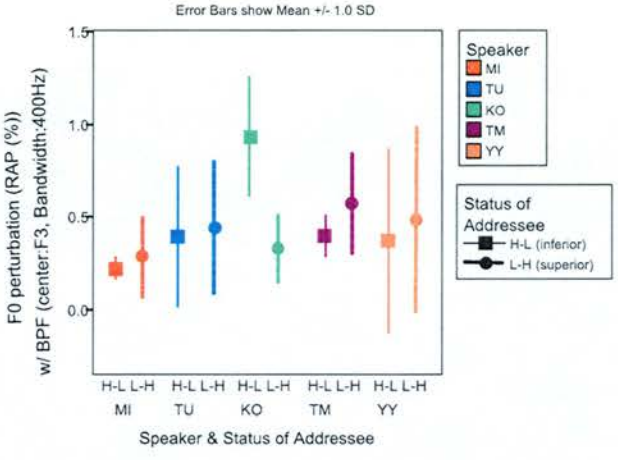
irregularity of the original waveform.

In contrast to Speaker KO, Speaker TM shows no association between the relative status of addressee and perturbations, but instead, he showed a strong and positive association between the relative status of addressee and Nw. However, for Speaker TM, no other acoustic parameters show significant association with the relative status of the addressee.

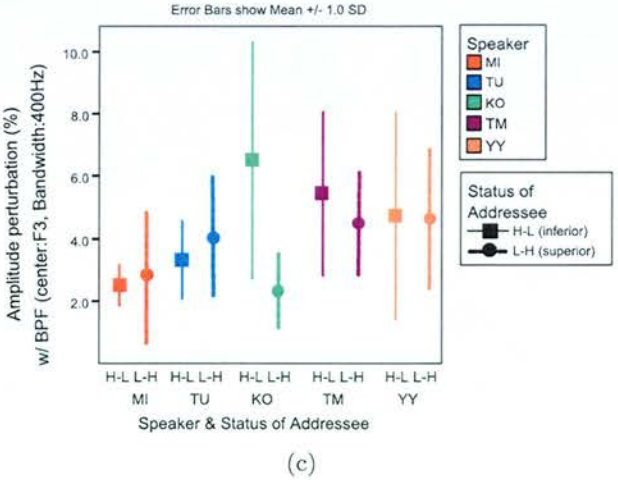
From these results, especially considering the result of Speaker KO, it seems plausible to suggest that some speakers may modify their voice quality according to the addressee's status, but not necessarily control the glottal configuration which could result in perceived breathiness.



(a)



(b)



(c)

Figure 4.10: The means (points) and standard deviations (lines) to compare the relative status of addressee with (a) noise ratings of waveform bandpass-filtered around $F3$ with a bandwidth of 400Hz (Nw), (b) F0 perturbation in the high frequency region (bandpass-filtered around $F3$ with a bandwidth of 400Hz), (c) amplitude perturbation in the high frequency region (bandpass-filtered around $F3$ with a bandwidth of 400Hz) across speakers.

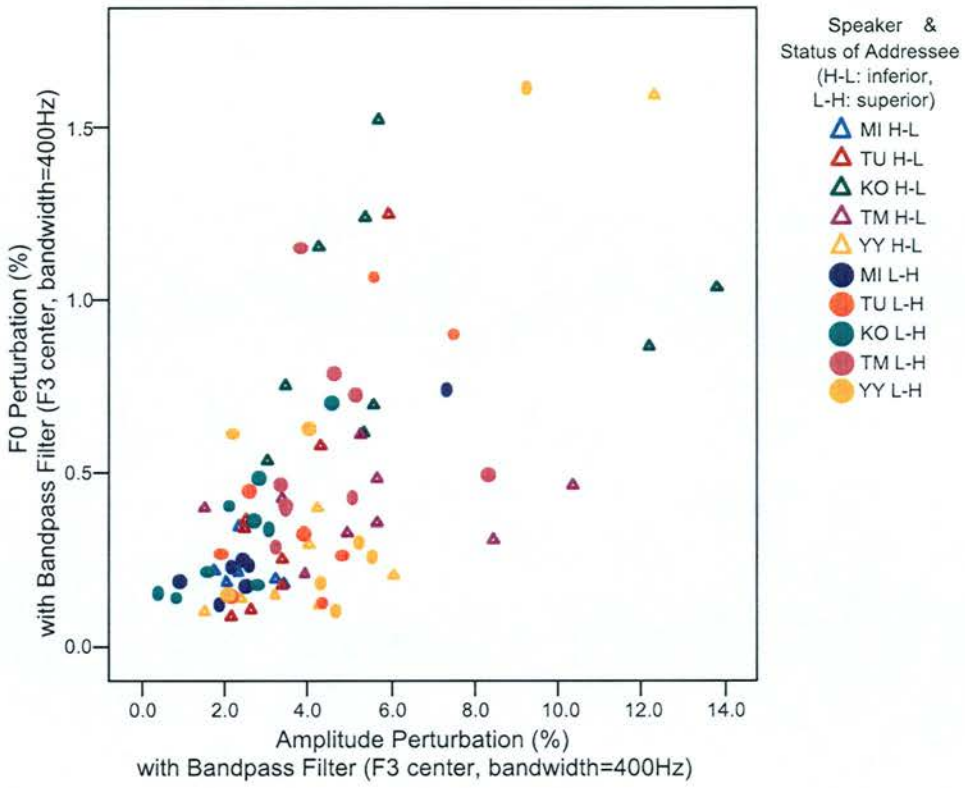


Figure 4.11: Relation between amplitude perturbation quotient after bandpass-filtered around $F3$ with a bandwidth of 400Hz and F0 perturbation quotient after bandpass-filtered around $F3$ with a bandwidth of 400Hz. Data points for the cases when the addressee is inferior are displayed as open triangles.

4.1.4.3 Comparison of spectral characteristics and noise-related parameters

Overall, spectral-parameters and noise-related parameters do not consistently show correlation with the relative status of the addressee. However, the following tendencies are found for each individual speaker.

Speaker MI $BW1$ has positive correlation with $H1^* - A1$ ($r = 0.634$, $p < 0.05$) and Nw ($r = 0.687$, $p < 0.01$). RAP_F3 and APQ3_F3 have a strong positive correlation with each other ($r = 0.868$, $p < 0.01$).

Speaker TU $H1^* - H2^*$ has relatively positive correlation with $BW1$ ($r = 0.602$, $p < 0.05$), and $H1^* - A1$ ($r = 0.725$, $p < 0.01$). RAP_F3 and APQ3_F3 have a positive correlation with each other ($r = 0.753$, $p < 0.01$).

Speaker KO $BW1$ and $H1^* - A1$ have a weak and positive correlation with each other ($r = 0.562$, $p < 0.05$). Nw has a weak correlation with RAP_F3 ($r = 0.497$, $p < 0.05$). RAP_F3 has a positive correlation with APQ3_F3($r = 0.606$, $p < 0.01$).

Speaker TM This speaker shows different tendencies from other speakers in various ways.

Nw has a good correlation with $H1^* - A3^*$ only ($r = 0.632$, $p < 0.01$). It is worth noting that this speaker has a larger $BW1$ than that of other speakers, however this speaker's $BW1$ does not correlate with other spectral characteristics of this speaker. The exception to this is the correlation between $BW1$ and $H1^* - A1$ ($r = 0.742$, $p < 0.01$), a possible indicator of $BW1$. From this result, Speaker TM's spectral characteristics are not dominated by $H1^*$ but may be dominated by one of the other factors (for example, $H2^*$ for $H1^* - H2^*$, A1 for $H1^* - A1$, and $A3^*$ for $H1^* - A3^*$). RAP_F3 and APQ3_F3 do not have any correlation with each other.

Speaker YY $BW1$ and $H1^* - A1$ have a moderate and positive correlation with each other ($r = 0.640$, $p > 0.01$) as predicted by theory and Nw has a positive correlation with $H1^* - A1$ ($r = 0.528$, $p < 0.05$). $H1^*$ is not a dominant factor of spectral characteristics for this speaker. RAP_F3 showed a positive correlation with Nw ($r = 0.580$, $p < 0.05$) and APQ3_F3 ($r = 0.811$, $p < 0.01$).

In addition, to check the validity of the Nw measurement, the presence of the second excitation in the waveform was examined according to the relative status of the addressee (Table 4.6). Only Speaker YY showed the second excitation consistently among tokens. This was especially evident when he produced his speech to his superior, but not to his inferior. However, this result is weak, because there were many tokens which could be interpreted as having the second excitation (i.e. Figure 4.5).

Table 4.6: Percentage of presence of the 2nd excitation in the waveform according to the relative status of the addressee (H-L: the addressee is the social inferior, L-H: the addressee is the social superior)

presence	speaker	MI		TU		KO		TM		YY	
of 2nd	status	H-L	L-H	H-L	L-H	H-L	L-H	H-L	L-H	H-L	L-H
excitation	no. of tokens	(6)	(7)	(8)	(8)	(9)	(9)	(9)	(10)	(8)	(8)
Positive (%)		0	0	12.5	12.5	11.1	0	0	11.1	37.5	87.5
Negative (%)		100	100	87.5	87.5	88.8	100	100	88.8	62.5	12.5

4.1.5 Production test: Summary

Overall, from the different associations between the spectral parameters which could be relevant to perceived breathiness and the relative social status of the addressee, we may say at least that the spectral parameters and the relative status of the addressee are not always associated in a distinctive manner with the same trend.

It seems that the usage of breathiness heavily depends on the voice characteristics or preferred strategy of the speakers. Regarding the result of noise ratings, only one speaker, Speaker TM, seemed to control his voice to express "general politeness" when speaking to his superior. Furthermore, the fact that Speaker TM showed a large BW1 corresponds with his noise rating results. However, the spectral parameters of Speaker TM did not show good correlation with noise ratings. The result of large BW1 and high Nw values contradicts the result that this speaker has moderate to small spectral parameters, which are not significantly different from those of Speaker KO. Therefore, the association between this speaker's glottal configuration and the relative status of the addressee was not confirmed.

On the other hand, Speaker KO's perturbation quotients and Speaker YY's second glottal excitation with his tense voice characteristics of $H1^* - A1$ and $H1^* - A3^*$ imply that the tension in vocal fold vibration is likely to depend on the relative status of the addressee. Therefore, we may reasonably suspect that their vocal fold vibration still might have resulted from physiological and psychological factors such as speakers becoming tense and nervous, which in turn leads to tension of the vocal folds. However, other physiological parameters such as vocal fold tension or heart pulse rate were not measured. Further discussions about this tension of the vocal folds will appear in the discussion chapter which considers the results of perception test.

4.2 Speech perception test: Responses to natural spontaneous speech

Here we come to further questions. Can listeners reasonably detect the speaker's intentional control or loss of control due to physiological change? Can both of these be perceived as "speaking to superior" and "polite"? Can physiological factors be obstacles for communication or can these extra-linguistic effects be considered by listeners?
To investigate the questions, a set of perception tests was conducted. The objective of the whole perception part of this project is to evaluate whether listeners can perceive "general politeness" related to the social relationship between speakers despite differences in individual speakers' voice quality characteristics and in individual speakers' vocal politeness strategies. The set of perception tests consisted of two parts: a judgement of lexical deference and a forced-choice listening test of the relative status of the addressee.

In a set of perception experiments, each token chosen for acoustic analysis was presented in a whole utterance without any manipulation. This allowed us to preserve the naturalness of the voice and avoid spoiling the acoustic quality by synthesis or a sudden interruption in an utterance by extracting a vowel of interest. The tokens were presented first in written form and later in speech so that it was possible to distinguish lexical politeness from politeness conveyed by vocal characteristics. This comparison could avoid the possible lexical influence from the use of *Keigo*. For example, some speakers use *Keigo* suffixes (i.e. *"-desu"*, *"-masu"*) quite often in some utterances, and listeners may perceive it as either

extra "formality", "deference", or speaker's idiosyncrasies. It is important to examine inter-speaker variation and listeners' preferences of *Keigo* use.

4.2.1 Lexical deference of utterances (script ratings)

4.2.1.1 Aim

In this study, the target tokens of /hidari/ were to be presented in the utterances in which they had appeared. When presenting in the utterances, there was a concern that listeners might get some lexical cues, such as formality by the use of "*teineigo*". The frequent use of "*teineigo*" might be misinterpreted as deference, since "*teineigo*" in present-day Japanese can express either formality or deference, especially when a speaker is not able to manage "*sonkeigo*" and "*kenzyoogo*". Even though the level of formality and deference of utterances was restricted by the control of participants' relationship and the situation, there was a need to eliminate this concern. Thus, lexical ratings were conducted prior to listening judgements. The purpose of this lexical rating was to elicit subtle differences in degrees of deference perceived on the basis of lexical cues.

4.2.1.2 Magnitude Estimation (ME)

In this study, Magnitude Estimation (ME) was employed for the rating method to measure the perceived degree of deference ("*kashikomari-do*") conveyed by lexical cues so as to confirm that the perceived degree of deference expressed lexically rated by stimuli were maintained in a certain range. Magnitude Estimation was originally introduced for measuring sensory impression in psychophysics (Stevens 1956, 1969), to associate the amount of physical magnitude (e.g. brightness of light, loudness of sound) with perceptual magnitude. In psycholinguistics, Bard et al. (1996) introduced ME to measure degree of linguistic acceptability.

Previous studies have mainly employed the linear scaling method and the comparative judgement with pair-wise presentations. Linear Scaling is a common rating method, using a shown scale and plotting a point. But there is a problem with expressing

distinctiveness between stimuli. For example, if we give the maximum number to one stimulus, and a later stimulus is stronger, there will be no available value to show distinctiveness. On a linear scale rating with a limited scope, scores given may not reflect the perceived degree of deference when several stimuli are at the end of the scale. With ME, there is no restriction of values in ratings, so the differences in perceived attributes can be expressed quantitatively, and become measurable.

Another common method that is used to investigate distinctiveness is the comparative judgement method, involving presentation of stimuli in pairs. But comparative judgement needs all the possible pair-wise combinations of stimuli to be presented. If subjects have limited time available for participating in experiments, this method is not realistic. Another problem with the comparative judgement method is that it does not allow for measurement of the degree of distinctiveness. By contrast, ME allows subjects to compare each stimulus with one modulus, and therefore the number of judgements for each subject will be reduced, compared with comparative judgement.

Despite the above mentioned advantages, one possible disadvantage of ME is that subjects are too used to Linear Scaling methods. Instructions need to be given carefully so as to allow subjects to rate the comparative magnitude of the stimuli and the modulus easily. Indeed, many subjects tended to respond on a linear scale, which was reported by Bard et al. (1996). To solve this problem, Bard et al. emphasised that the most important thing is to give careful instruction before the experiments. In this study, instruction about scoring the magnitude was given as follows. Raters were asked to give a number greater than one if a stimulus is more deferential than the modulus, or to give a number between zero and one if a stimulus was perceived to be less deferential than the modulus, so as to convey a multiple fraction. For example, if the stimulus was perceived twice as submissive as the modulus, then it would receive a rating of 2, while if the stimulus was perceived half as submissive as the modulus, then it would receive a rating of 0.5 (=1/2).
Even so, it is quite natural for many speakers to respond on a linear scale because it is familiar to them. Some subjects may respond on a power scale whereas others may

respond on a linear scale, which does not allow an experimenter to compare subjects in a straightforward way. Therefore, normalisation according to each subject's responses was applied in this study. If responses from a rater followed the normal distribution, then the responses would be treated as though they were in the linear scale and therefore would be normalised to the score ranged between 0 and 1 without taking the logarithm. On the other hand, if responses from a rater followed the power scale, then the rater responded as we intended, and after taking the logarithm the responses were normalised to the same score ranged between 0 and 1. This normalisation was computed for each rater because the preference of the scale depends on each rater.

4.2.1.3 Subjects

To avoid the effects of dialect and regional cultural background on prosodic features, native speakers of Tokyo dialect were recruited to participate in this experiment. A total of twenty-two people (eleven males and eleven females) participated in this experiment. Fifteen of the subjects were university students without any work experience, and the other seven subjects were university graduates with work experience. The majority of them were born and brought up in the Tokyo area, and all participants had been residents in the Tokyo area for more than four years.

4.2.1.4 Materials and Procedure

Subjects were tested individually. They were first given a practice model to follow. They were then presented with a set of stimuli from the collection of materials after a modulus (which was to be used as a standard reference of deference in this experiment). ME is supposed to give no restriction in rated values given in a ratio scale, so the responses of each rater should be maintained in a logarithmic scale. However, a modulus utterance was chosen for each set subjectively by an experimenter, so that the modulus is likely to express neutral deference level in each set. Raters were required to estimate the magnitude of deference of each stimulus, as explained in the earlier section. Each input field was displayed to the participants on a PC, together with a stimulus sentence one by one, and the subjects were asked to give their estimated score. A set of stimuli, consisting of

thirteen to nineteen phrases including the word /hidari/ from each speaker, was randomly mixed with dummy utterances. The dummy utterances were inserted so as to distract the raters' attention from the word /hidari/ which contained the target vowel. Each session started with a speaker's modulus followed by a set of these 32 stimuli. The subjects were presented with one set for each of the five speakers. Thus a total of 160 stimuli, including 82 tokens of /hidari/ were presented. As suggested in the discussion of ME above, responses were processed and normalised by subject, depending on each subject's preference of power/linear scale, so that raw scores were converted to a linear scale, ranging between 0 and 1 for each subject.

4.2.1.5 Result

Figure 4.12 shows the mean and standard deviation of normalised ME ratings of lexical cues across the speakers and the relative status of addressee. The results shown in the diagram indicate that Speaker TU might have successfully conveyed his deference through lexical cues. In other words, lexically neutral utterances might have not been selected for this speaker, which should be considered in the following auditory experiment. However, considering the overlapped ranges of standard deviation across the relative status of the addressee, most of the stimuli could not be distinguished clearly by lexical cues, therefore they might be still perceived as either addressing superior or inferior. Given the fact that the degree of lexical deference seems to be indistinguishable in the provided stimuli, we will be able to observe the effect of vocal cues in conveying the politeness to superior by the perception tests with the stimuli.

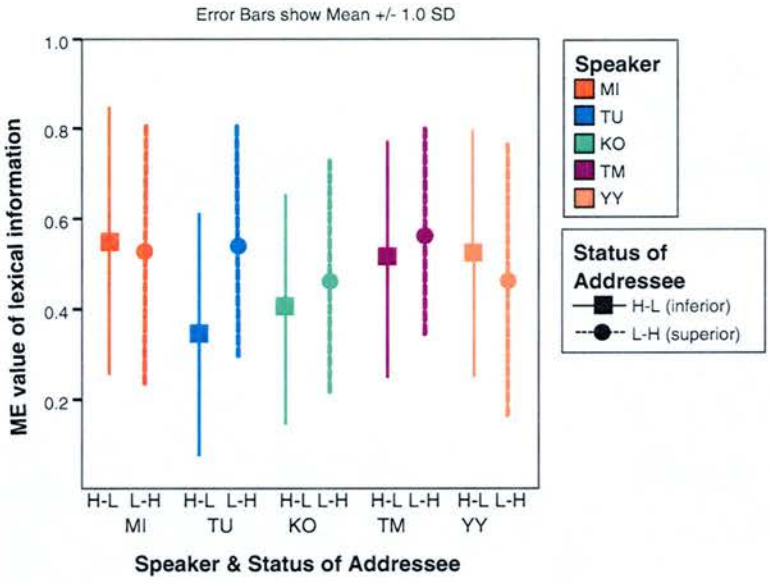


Figure 4.12: The mean and standard deviation of magnitude estimation ratings of lexical deference across the speakers and the relative status of the addressee.

4.2.2 Speech perception test: forced choice

4.2.2.1 Aim

To determine whether Japanese speakers can perceive the relative status of the speaker and the addressee based on the voice quality of a speaker, we used a forced-choice judgement task based on utterances taken from the recordings of the Map Task. This judgement gives a possible measure of the influence of the "general politeness" to superior signalled in speech. In a pilot study (Ito 2002), many subjects expressed having difficulty in the ME ratings of deference when they were asked to rate audio stimuli. This experiment therefore aimed at eliciting more natural responses, such as judging if an utterance sounds appropriate for use in addressing one's superior or inferior, which happens in our daily life in Japan. Thus, a forced-choice procedure was selected for this task. First, the result of this listening experiment, the perceived degree of "general politeness to superior" was compared with the result of the ME ratings of lexical deference. After considering the effect of lexical deference to the general politeness to superior, the result of this experiment is compared with acoustic parameters measured in the production experiment, which could be associated with perceived breathiness.

4.2.2.2 Subjects

All the subjects who had participated in the lexical deference judgement subsequently participated in this task.

4.2.2.3 Materials and Procedure

Subjects were tested individually. They were presented with a set of stimuli through headphones, utterance by utterance, from the collection of materials. After each utterance, they were asked to choose the relative status of the addressee of each stimulus, which was either superior or inferior. Each input field was displayed on the PC, together with a stimulus utterance played from the subjects' headphones. A set of stimuli consisted of thirteen to nineteen phrases including the word /hidari/ from each speaker which were employed for acoustic analysis and lexical judgement. These were randomly mixed with

dummy utterances, to keep raters' attention away from the target word. Thus, in the course of each listening session, the listeners were presented with a set of these 32 stimuli consisting of target tokens and dummy utterances for each of the five speakers. A total of 160 stimuli, including 82 tokens of /hidari/ were presented.

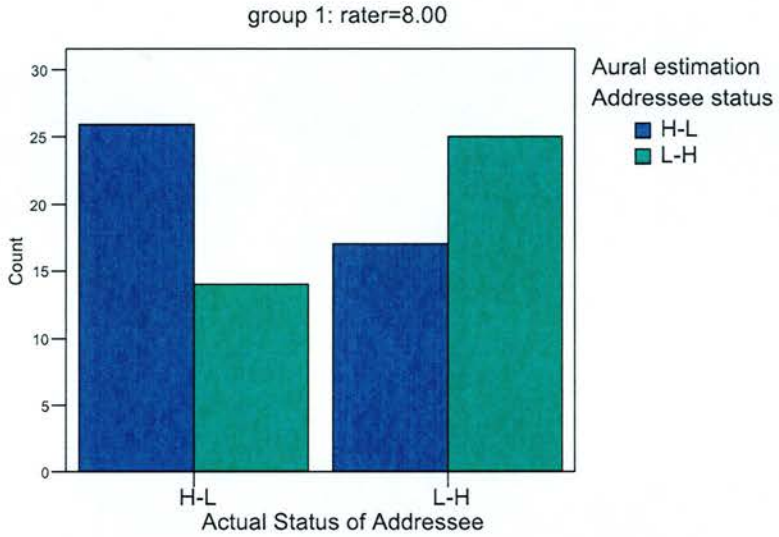
4.2.3 Results of perception tests

4.2.3.1 Strategies preferred in detecting politeness to a social superior

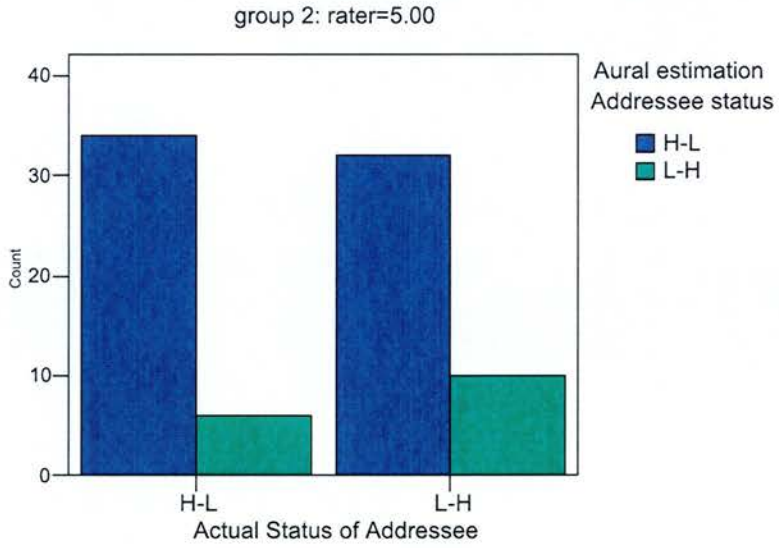
To check whether the raters could use acoustic cues effectively, the data were analysed by using a binomial test. From this analysis, two groups of subjects were found. Thirteen out of the twenty-two raters judged the relative status of addressee correctly in the majority of utterances, with a significant difference ($p < 0.05$) (e.g. Figure 4.13 (a)). Eight out of the twenty-two raters judged that the majority of utterances sounded as though they were addressing the speakers' social inferior, without a significant difference (e.g. Figure 4.13 (b)). One rater's responses did not correspond to either of these two groups, so this speaker's responses are excluded from further analysis. In the following analyses, we consider the two main subject groups separately.

In addition, we consider the influence of speaker characteristics on raters' responses in judging the relative status of the addressee – that is, we attempt to see if some speakers are likely to be judged consistently as addressing a superior or an inferior (Figure 4.14). From these results, we could see that both groups of raters consistently tended to perceive utterances spoken by Speaker MI (Figure 4.14 (a)) and Speaker KO (Figure 4.14 (c)) as addressed a social inferior. For Speaker MI and KO, that is, aural presentation made it more difficult to judge the addressee's relative status correctly. For the stimuli from other speakers, there was no such overall tendency, and aural presentation helped in judging the relative addressee status correctly.

Comparing the results from Speaker TU (Figure 4.14 (b)), Speaker TM (Figure 4.14 (d)), and Speaker YY (Figure 4.14 (e)), who were perceived to be addressing a social superior correctly by group 1, it is not always *Keigo* which works in detecting the general politeness shown to a higher status addressee, when acoustic cues are available. Speaker TU showed



(a) group1



(b) group2

Figure 4.13: Examples of two different types of responses obtained from raters. Group 1 (a) judged the status of the addressee relatively correctly whereas Group 2 (b) rated most of stimuli as to social inferior. The horizontal axis shows the actual relative status of the addressee and the vertical axis shows the observed number of samples. The blue bar in each cluster shows the frequency of samples that rater judged as addressed the social inferior and the green bar in the right side in each cluster shows the frequency of samples that the rater judged as addressed the social superior.

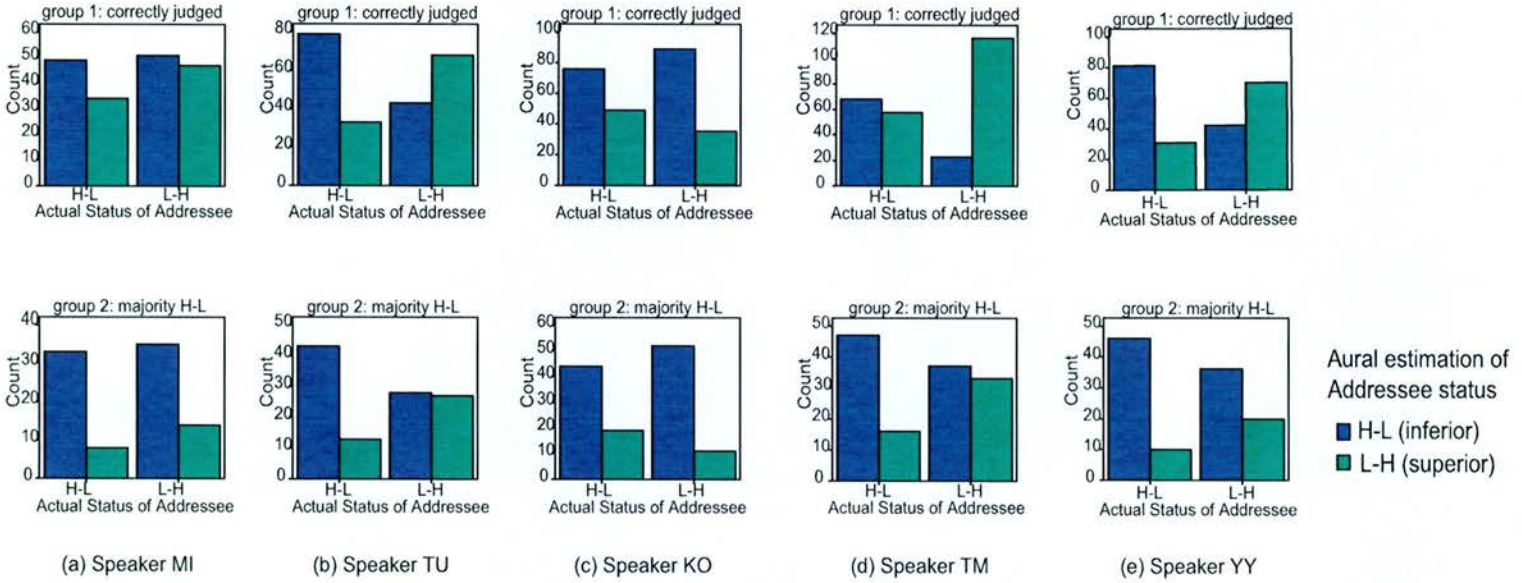


Figure 4.14: Responses to the audio stimuli of each speaker obtained from raters. (Group 1: raters who judged the status of the addressee relatively correctly. Group 2: raters who judged most of the stimuli as produced by social superior.) The **horizontal** axis shows the **actual** relative status of the addressee in the Map Task dialogues, the **colours** represent the **perceived** relative status of the addressee by raters, and the vertical axis shows the observed number of samples. The blue bar in each cluster shows the frequency of samples that the raters judged as addressed the social inferior and the green bar in each cluster shows the frequency of samples that the raters judged as addressed the social superior. The left cluster in each diagram shows the number of the tokens actually addressed the social inferior and the right cluster shows the number of tokens actually addressed the social superior.

the best perceived deference with his lexical cues (Figure 4.12), whereas Speaker TM showed the best perceived politeness to superior with audio stimuli.

4.2.3.2 Listening judgement results: Correlation between two groups

Even though the two groups of subjects responded differently (Group 1: tended to judge the relative status of the addressee correctly on the basis of vocal cues, and Group 2: tended to judge the relative status of the addressee as a "social inferior"), we need to examine how close or different their judgements are, to consider the effect of vocal and lexical cues. To examine this, speech scores (S scores) which represent the degree of general politeness to superior, using both lexical and vocal cues, were computed as follows.
For each group, the proportion of raters who rated the addressee as a social superior of the speaker was computed for each stimulus, and the proportion was used as S score. Correlation between the S scores obtained from the two groups of subjects was computed for each speaker. The S scores from group 1 and group 2 for Speakers TU, TM, and YY were significantly correlated ($r = 0.862$, 0.581, and 0.811, respectively, all $p < 0.01$), even though group 2 tended to rate these speakers' voice as not addressing a superior, when a superior was actually addressed. The S scores from group 1 and group 2 for Speakers MI and KO were not correlated.

4.2.3.3 Influence of lexical cues on listening judgement: result

To check the degree of influence of lexical cues on listening judgements of the relative status of the addressee per speaker per group, the correlation between the mean of ratings of Magnitude Estimation (ME ratings) of lexical cues and the S scores per stimuli was computed (Table 4.7).

From this table, it can be seen that the perceived deference conveyed by lexical cues and the perceived general politeness to superior conveyed by the voice may be quite different, especially in the case of Speaker MI whose ME rating score (lexical deference) and S score (vocal politeness to superior) are strongly negatively correlated. *Keigo* may be used to judge deference, but it did not seem to affect the raters' perception of the relative status of the addressee when they were exposed to audio stimuli of Speaker MI. There

Table 4.7: Pearson product moment correlation coefficients between the mean of normalised value of magnitude estimation rating of lexical stimuli and the S score obtained from the group of raters. The coefficients were computed per each speaker. (group 1: tended to judge the correct status of addressee, group 2: tended to judge the addressee as a social inferior). (Correlation is significant at *: 0.05, **: 0.01 level (2-tailed).)

	speaker	S score of listeners group	
		group1 (judged correctly)	group2 (judged as inferior)
ME lexical ratings mean	MI (N=13)	**-.758(**)**	-.483
	TU (N=16)	-.218	-.109
	KO (N=18)	.076	**.495(*)**
	TM (N=19)	.150	.312
	YY (N=16)	.089	-.019

was a strong negative correlation between the S score and lexical ME ratings of group 1 for Speaker MI. On the other hand, there was a moderate positive correlation between the S score and the lexical ME ratings of group 2 for Speaker KO. These two are the only cases where the lexical cues may have interfered with listeners' judgements of the relative status of the addressee, and the two cases go in opposite directions. Overall then, the results give us reason to believe that the judgements of general politeness based on auditory presentation are a valid reflection of something other than lexical deference.

4.2.3.4 Acoustic parameters vs. listening judgement results (Speech Score)

To investigate which acoustic parameters analysed in the production test are influential in judging the relative status of the addressee, the correlation between acoustic parameters obtained in the production experiment and the S score averaged across listeners for each group was computed (Table 4.8). Scatter-plot diagrams also show the correlation between spectral parameters and the S score (Figures 4.15- 4.19).

Table 4.8: Pearson product moment correlation coefficients between acoustic parameters and the S score obtained from the group of raters. The coefficients were computed per each speaker. "gp" stands for group of raters, (1: tended to judge the correct status of addressee, 2: tended to judge the addressee as a social inferior).(Correlation is significant at *: 0.05, **: 0.01 level (2-tailed).)

speaker	gp.	$BW1$ (Hz)	H1*-H2* (dB)	H1*-A1 (dB)	H1*-A3* (dB)	Nw	RAP_F3 (%)	APQ3_F3 (%)	RAP_F3 −RAP(%)	APQ3_F3 −APQ3(%)	RAP (%)	APQ3 (%)
MI	1	-.615 (*)	-.483	-.167	-.556 (*)	-.405	.102	.277	-.222	-.162	.082	.035
(N=13)	2	-.386	-.463	-.288	-.370	-.315	.225	.032	.058	.109	-.019	.124
TU	1	-.068	.176	-.204	-.478	.475	-.089	.019	-.149	-.550(*)	-.152	-.396
(N=16)	2	.042	.102	-.128	-.257	.347	-.053	.114	-.081	-.341	-.011	-.128
KO	1	-.075	-.087	-.247	-.300	-.337	-.109	-.193	.359	.381	-.137	-.175
(N=18)	2	-.385	.059	-.172	-.244	.134	.291	.313	-.008	-.242	.183	.002
TM	1	.052	.093	.174	.264	.418	.403	-.325	.422	.476(*)	.207	.208
(N=19)	2	-.224	.211	.000	.533(*)	.535(*)	.365	-.164	.088	.156	.019	-.070
YY	1	-.659 (**)	-.095	-.409	.363	-.171	.134	.141	-.415	.157	-.205	.342
(N=16)	2	-.561 (*)	-.128	-.279	.402	-.018	.127	.321	-.354	-.164	-.012	.280

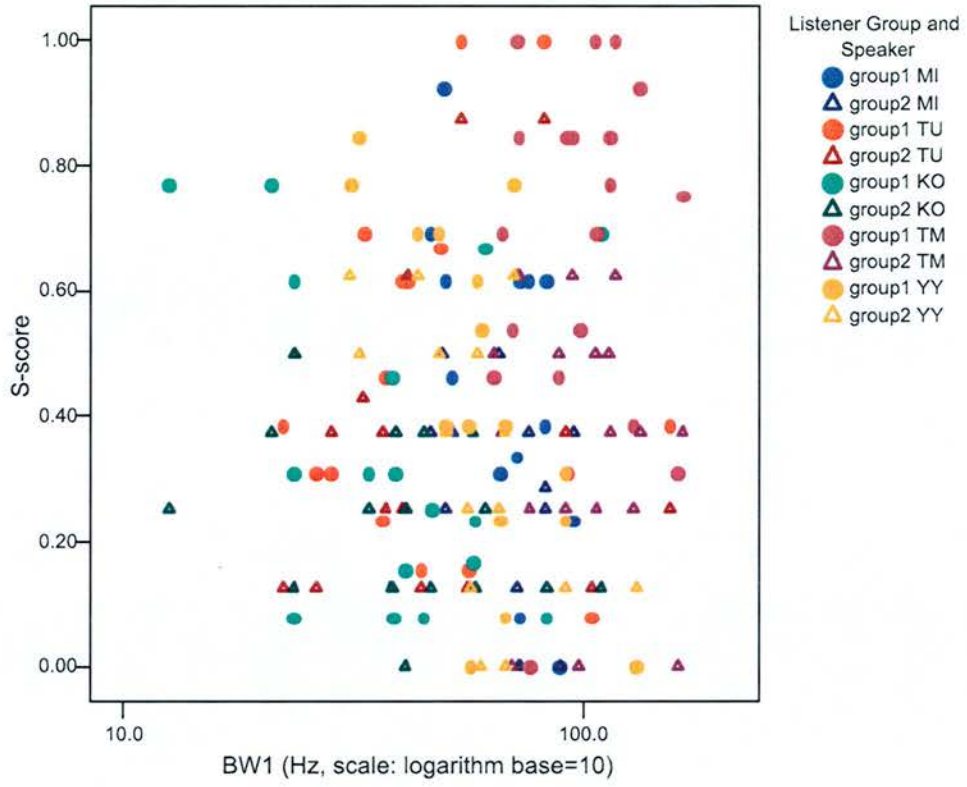


Figure 4.15: Relation between $BW1$(Hz) and the S score (the rate of raters who perceived the stimuli as addressed a social superior). Data points for the cases of the S score from rater group 2, who were likely to perceive that the addressee was inferior, are displayed as open triangles.

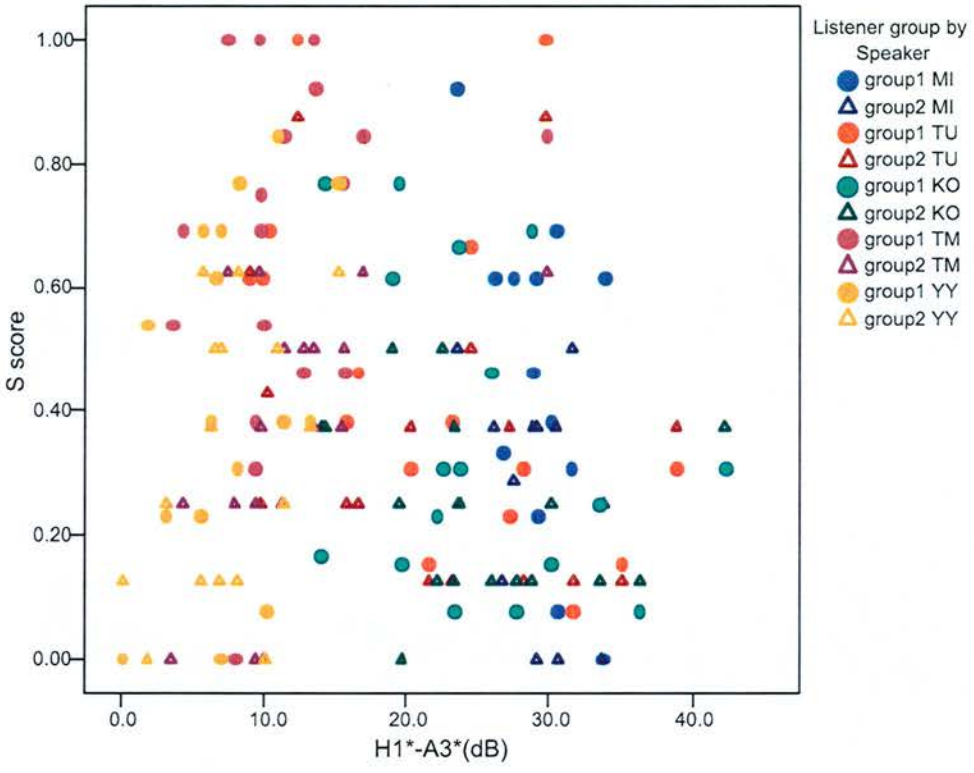


Figure 4.16: Relation between $H1^* - A3^*$(dB), a possible indicator of spectral tilt, and the S score (the rate of raters who perceived the stimuli as addressed a social superior). Data points for the cases of the S score from rater group 2, who were likely to perceive that the addressee was inferior, are displayed as open triangles.

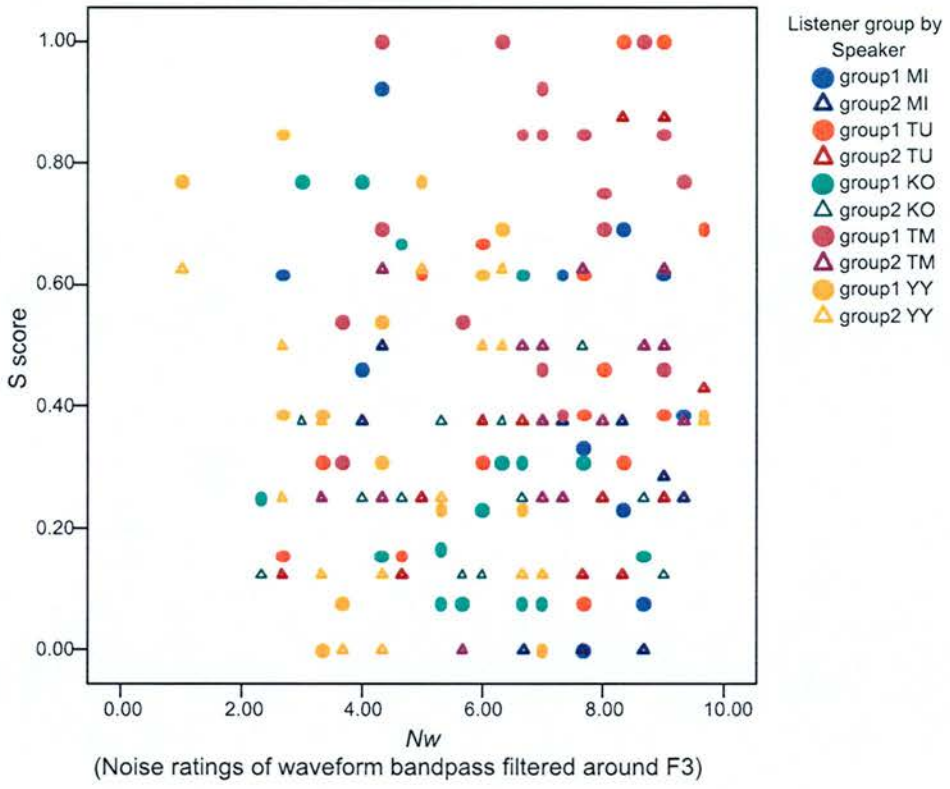


Figure 4.17: Relation between the S score (the rate of raters who perceived the stimuli as addressed a social superior) and Nw (noise ratings of the waveform bandpass-filtered around F3), a possible indicator of aspiration noise. Data points for the cases of the S score from the rater group 2, who were likely to perceive that the addressee was inferior, are displayed as open triangles.

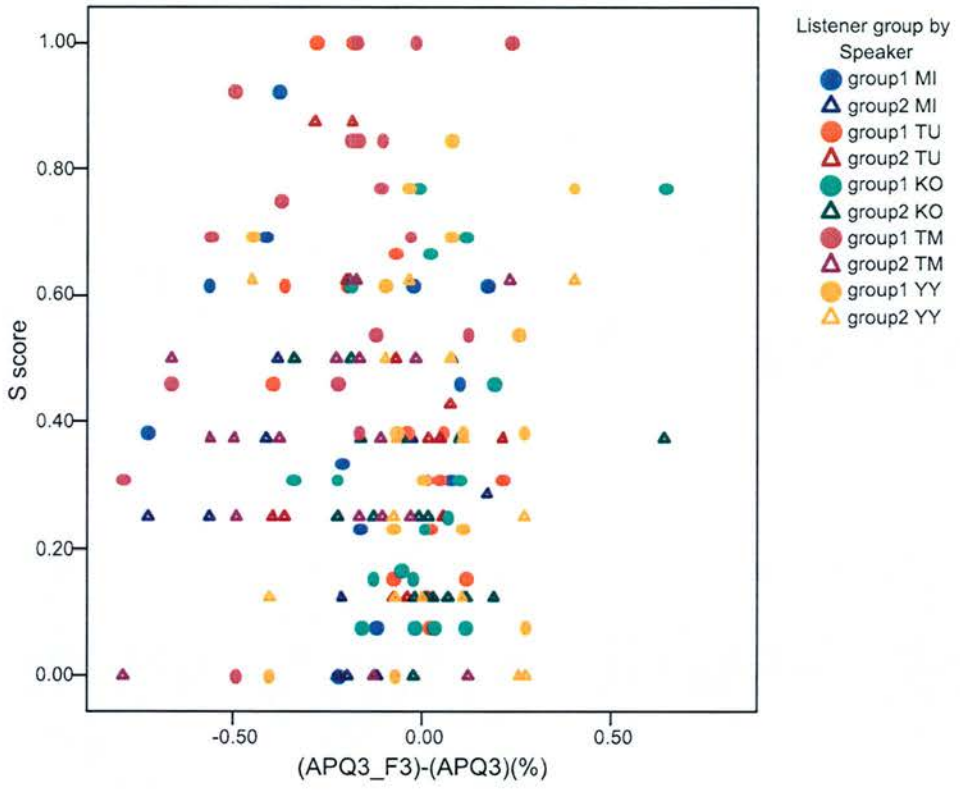


Figure 4.18: Relation between the S score (the rate of raters who perceived the stimuli as addressed a social superior) and the difference between the amplitude perturbation quotients of the original waveform and the bandpass-filtered waveform around the F3 region. Data points for the cases of the S score from rater group 2, who were likely to perceive that the addressee was inferior, are displayed as open triangles.

First, I report the result of tendencies observed for each group(=gp) of listeners.

Gp1: listeners who tended to rate the relative status of the addressee correctly

This group has negative correlation between their S scores and both $BW1$ and $H1^* - A3^*$ of Speaker MI. This group's S score also has correlation with the relative amplitude perturbation quotients, APQ3_F3-APQ3. When this group listened to Speaker TU and Speaker TM, however, their reaction was the opposite: for Speaker TU, the association is negative, whereas the association is positive for Speaker TM.

Gp2: listeners who tended to rate the relative status of the addressee as a social inferior

This group tends to have positive correlation between their S score and $H1^* - A3^*$ and Nw, when they hear the voice of Speaker TM.

Second, I describe some trends observed per speaker.

Speaker YY Both groups have negative correlation between their S score and $BW1$.

Speaker KO The irregularity in F0 and amplitude, which Speaker KO changed according to the actual relative status of the addressee, did not seem to assist listeners in their perception of the relative status of the addressee. None of the acoustic parameters seemed to affect judgements of the relative status of the address for this speaker, which corresponds with the fact that all the raters in both groups had tended to judge the relative status of the addressee incorrectly with this speaker's voice.

Third, I discuss some overall findings across speakers and listeners groups.

Interestingly, the direction of association between the $H1^* - A3^*$ and the S scores seems to change according to the individual speaker's range of $H1^* - A3^*$, which is a potential indicator of spectral tilt. Speakers who are at the low end of the range of spectral tilt tend to have a positive association between spectral tilt and the S score (refer to Figuré 4.16, Figure 4.19 and column $H1^* - A3^*$ of Table 4.8). On the other hand, other speakers who have relatively high spectral tilt values tend to have a negative association between spectral tilt and the S score.

If this is the case, then it might be possible that listeners change their strategy for interpreting spectral tilt according to the range of each speaker's spectral tilt. At first glance, this result might look contradictory across speakers. However, if we examine this result in the light of the argument put forward by Usami (2002), we can find some consistency in the usage of spectral tilt for judging the relative status of the addressee. Usami stated that people avoid extremity when expressing politeness. If the listeners assumed the extreme ranges of spectral tilt as signalling "*im*politeness", then they naturally assumed that the speakers would use a neutral spectral tilt range when they speak to their social superior. Therefore, the change in the usage of spectral tilt could be treated as a possible acoustic cue for perceiving politeness to a social superior.

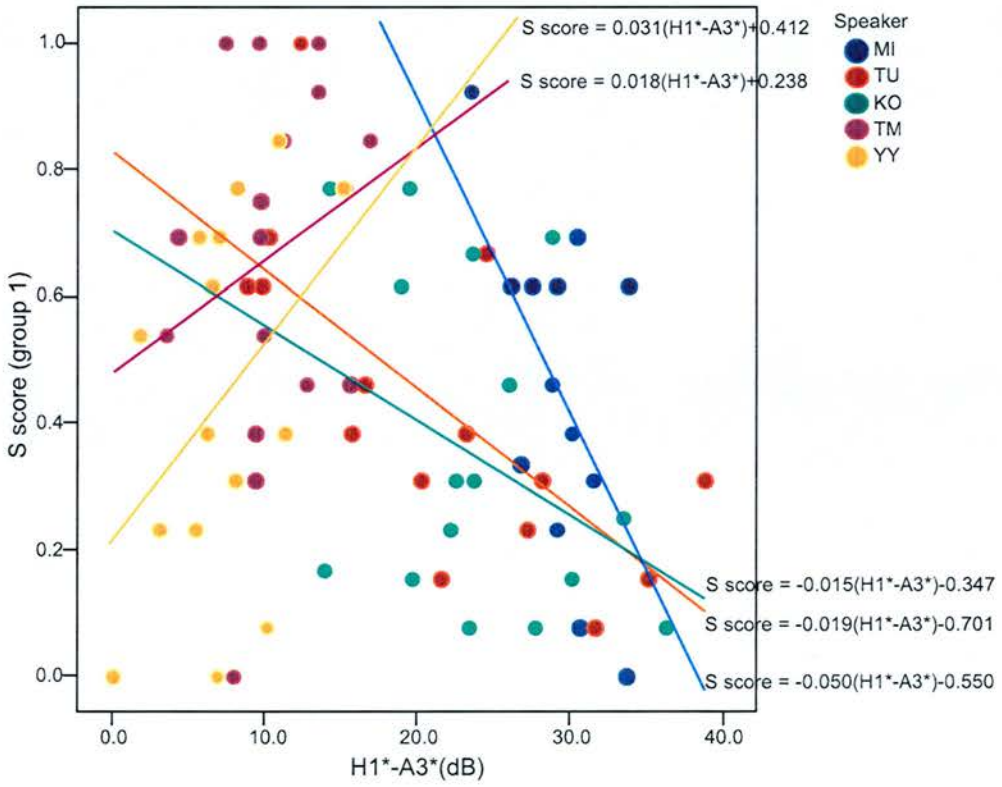


Figure 4.19: Relation between $H1^* - A3^*$(dB), a possible indicator of spectral tilt, and group 1's S score (the rate of raters who perceived the stimuli as addressed a social superior). Regression lines represent the linear regression between $H1^* - A3^*$ and the S score of group 1 per speaker.

From the results above, it appears that increased amplitude perturbation quotient increased in the $F3$ region may be involved in judging the relative status of the addressee correctly, and that the range of spectral tilt per speaker seems to contribute to listeners' choice of strategies in perceiving the general politeness to superior. However, it is difficult to observe consistency in the way these acoustic parameters are used across speakers and across listener groups. We cannot identify one modification of voice quality that has the most effect on changing the impression of the politeness expressed to a social superior.

4.2.3.5 Perception of politeness and breathiness: Summary

The results of the perception experiments showed the following tendencies.

First, at least some raters successfully used vocal paralinguistic features for judging the relative social relationship between the speakers. This is especially true for the majority of raters who tended to judge the actual relative status of the addressee correctly when they were exposed to the audio stimuli. Relying on lexical information conveyed by *Keigo* alone cannot present an effective strategy to make the right guess of the social status of the addressee.

Second, some listeners tended to judge that utterances sounded as though they were addressed social inferiors than social superiors. One possible explanation for this behaviour is that the voice simultaneously conveys positive politeness (which might shorten the social and psychological distance) and the negative politeness of deference. Alternatively, the effect of voice may contradict the effect of *Keigo*, that is, to keep distance between a speaker and an addressee. Either way, this contradiction between expression of negative and positive politeness may have led these listeners to misjudge the relative status of the addressee.

Third, the increase of amplitude perturbation quotient in the high frequency region may be involved in the perception of the politeness to superior. The most interesting fact here is that the range of spectral tilt per speaker seems to be used by listeners to modify their strategy for estimating how speakers' voice quality is affected by the relative status of the addressee. When a speaker has low values of spectral tilt, then a higher spectral tilt

value tends to be perceived more polite, whereas a speaker with high values of spectral tilt tends to be perceived as more polite when the spectral tilt value is lower. This change in the usage of spectral tilt as an acoustic cue for perceiving the politeness to superior suggests that listeners tend to perceive the middle range of spectral tilt across speakers as appropriate for speaking to a social superior. Considering Usami's argument (2002) that people avoid extremity when expressing politeness, this neutral spectral tilt range could be the cue to express their politeness to superior.

In summary, it is possible that when perceiving politeness to a superior addressee, listeners use as a cue some sort of voice quality (i.e. breathiness which is related to a certain glottal configuration). However, the parameters used to express/perceive politeness vary from speaker to speaker and from one group of listeners to the other. Some parameters did not seem to be employed by the listeners with consistency and it is difficult to assert that the listeners used the degree of breathiness as a cue for politeness. Even so, the fact that the listeners seem to have used spectral tilt as a cue for judging the addresses' status confirms that there is certain contribution from voice quality to conveying politeness.

Chapter 5

Discussion

In this chapter, I summarise the significance of the findings of this thesis and I discuss the problems which remain for the future. First, the experimental design of the procedure for eliciting spoken politeness is examined. Second, the acoustic analysis of speech production is reviewed. Subsequently, the results of the speech perception tests of politeness are examined. Finally, the general problems of studying vocal paralinguistics are discussed and some solutions deduced from this study are suggested.

5.1 Experimental design of eliciting politeness: speech data collection

As discussed in Chapter 2, there are several strategies that can be used for expressing politeness. According to Brown and Levinson (1989), the strategy taken would depend on the situation, such as the relationship between speakers, and the context of the conversation. Therefore, it is necessary to control the situation in which speech samples are collected.

The methodology of the Map Task, which is employed in this study, enabled us to exert a fair degree of control over these factors. Specifically, all of the following were well controlled: the relative status of the addressees (which determines the degree of deference to be expressed by the speakers); the familiarity between participants (which heavily influences the degree of formality); and the phonological environment of the utterances chosen

for analysis (which reduces the variability of the material analysed acoustically). A number of researchers, especially Western non-sociolinguists (e.g. Nick Campbell, personal communication), have expressed doubts about the validity of the Map Task for eliciting politeness. They argue that excess control might increase the degree of formality which comes from the awareness of being recorded, and thus might obscure the degree of formality and deference which comes from the status difference. However, from the results of the frequency analysis (Appendix A) of *teineigo* suffixes (*-desu, -masu, -kudasai*), it is clear that none of the speakers ever violated the *Keigo* functions in their utterances. This fact supports the predictions of Shibatani (1990) and Jorden (1962) that people use *teineigo* as a part of grammar, depending on the formality of the situation and familiarity between speakers, but not from politeness, especially deference. In our Map Task recordings, speakers followed the *Keigo* rule, where they are supposed to use *teineigo* at all times when they speak to their superior, but not necessarily when they speak to their inferior. It is reasonable to say that the Map Task successfully elicited unscripted, semi-natural, spontaneous speech.

At the same time, the results of the frequency analysis of *teineigo* also indicate that the function of *Keigo* in expressing politeness has been exaggerated by traditional researchers of the Japanese language and should be re-examined. The fact that some speakers kept the degree of formality by *Keigo* when they spoke to both their superior and inferior, but seldom used *sonkeigo* (respect form) and *kenzyoogo* (self-humbling form) suggests that *Keigo*'s main function may have shifted from expressing deference to expressing formality only, as has been reported by Inoue (1989, 1998).

5.2 Speech production and acoustic analysis

In this study, measurement of glottal characteristics associated with perceived breathiness was mainly done directly from the waveform and spectrum, as suggested by Klatt and Klatt (1990) and Hanson (1995). For studying natural spontaneous speech, the direct waveform and spectrum measurement has the advantage of being non-invasive. However, in the case of male speakers, their voice may contain a second glottal excitation which appears in the waveform and interferes with the result of direct waveform and

spectrum measurement (Hanson 1999). Moreover, when I presented my earlier findings at conferences and workshops, some researchers questioned the reliability of subjective visual observation used in the noise rating, even though the high correlation between responses of raters seems to guarantee the validity of the obtained data. Therefore, in addition to the direct waveform and spectrum measurement, the following two analyses were conducted: 1) the observation of the second glottal excitation, 2) the measurement of F0 and amplitude perturbation quotients in the high frequency region as an alternative to noise rating. The results show that three out of the five speakers seem to show some changes in their voice according to their addressee's relative social status. Therefore it is possible to say that speakers may change not only the level of formality of *Keigo*, but also their voice, when expressing formality, deference or solidarity. On the other hand, the results also indicate that the acoustic characteristics which changed according to the addressee varied from speaker to speaker, such as shimmer of the irregular vocal fold vibration, second glottal excitation from overall tension of vocal folds, and noise rating. The difference of acoustic characteristics above is due to the behaviour of vocal folds, such as tension at the vocal folds or posterior chink. It is reasonable to think that when speakers address someone of a different status, they change the manner of their vocal fold movement, thus glottal characteristics are still the focus of interest.

Overall, the speakers do change their voice quality according to the addressee. Even under this controlled situation, a great variety of voice qualities were produced. There are many factors to be considered, not only sociological, but other factors such as physiological and psychological. From this small number of speakers, it is impossible to say whether this difference comes from social and cultural preferences, personality, or physiological constraints.

To explore these factors in changing vocal fold settings in subsequent work, the following two approaches should be considered. First, a larger number of speakers should be studied to ensure that trends noticed above are also present in a larger and more diverse group of speakers. If we could pool data from a larger number of speakers, we could

possibly determine: 1) whether the range and the change of trend in spectral tilt depend on the speaker's physiologically available range of spectral tilt or merely the preference of a speaker, and 2) which trend of spectral tilt changes, the increase or decrease, would be more popular among Japanese male speakers.

Second, another method of exploring glottal characteristics will be needed. For example, the data could be analysed by source-filter decomposition, as suggested by Gobl (1989). By the analysis of source-filter decomposition, we may capture the proportion of contributions from either vocal folds (source) and vocal tract (filter) in the change of spectral tilt. However, this requires a certain level of experience and skill, and is a time consuming process.

5.3 Perceived politeness in speech

The forced-choice of addresse status test in contrast with the ME lexical deference test revealed the following facts.

1) The majority of listeners tend to use vocal paralinguistic cues to judge the general politeness to superior. They may also take into consideration the appropriateness of the degree of formality and deference expressed by *Keigo*; however, in ordinary conversations, when all the participants are MPs (model persons) who are capable of using *Keigo* properly, speakers never violate the appropriate *Keigo* usage in expressing formality and deference. Therefore, in their judgement of the politeness to superior, listeners need to rely heavily on vocal paralinguistics. Furthermore, listeners who tend to rely primarily on lexical cues often fail to estimate the level of deference, although speakers change their voice according to the relative status of the addressees. Not all the speakers managed to communicate the politeness to superior they were supposed to express. Therefore, even though there was no violation of *Keigo* usage, many listeners tended to perceive the majority of speech tokens as less polite than was supposed to be expressed by the speaker to superior. The fact that deference-forms were very rare in the collected speech data also supports the idea that the function of *Keigo* in expressing politeness, especially deference, has been overestimated by traditional researchers on Japanese.

In studying spoken conversation, therefore, voice characteristics should be studied carefully, as long as the lexical content satisfies the appropriate level of politeness.

2) Listeners' judgements of the relative social status of the addressees are moderately correlated with two parameters, both in the third formant ($F3$) region. These were mainly spectral tilt ($H1^* - A3$) and amplitude perturbation around $F3$. From this result, it seems that the listeners actively employed the information conveyed in the high-frequency region. If we consider that aspiration noise is prominent in the higher frequency region, as suggested by Hanson (1995, 1997, 1999), the fact that listeners' judgements correlated with amplitude and F0 perturbation around the $F3$ region suggests that breathiness must be involved in judging politeness. Furthermore, compared to the noise rating method, the quantitative method gives us more reliable measures, which led to the result that the significant correlations between irregularity of the waveform in the $F3$ region and politeness judgements were found in more speakers than were found using the noise rating method.

In this study, breathiness rating was not conducted because of the limited number of speakers and listeners. Therefore, it is difficult to say whether or not listeners' responses were influenced by breathiness in the tokens. In spite of this, it is possible to say that there was a tendency across speakers to have significant correlations between the irregularity of the waveform in the $F3$ region when measured in a quantitative manner, and the judgement of relative status of addressee. This tendency supports the idea that acoustic cues in the $F3$ region are used.

This is a subject for future work. Perception studies on glottal characteristics of the voice will make it possible to distinguish the perceptual contributions of acoustic characteristics such as perturbations of amplitude and frequency, signal to noise ratio, and resonances (formants).

For example, a synthesiser which enables us to control the acoustic characteristics above (Antoñanzas-Barroso et al. 2005) could be used experimentally to demonstrate the contribution of these characteristics to perceptible voice quality difference in the $F3$ region. To examine these characteristics, signal processing techniques designed to detect

them (de Krom 1994) are necessary.

3) Different speakers used different strategies of voice quality changes for conveying "general politeness". Despite these differences, some listeners who relied on vocal cues were able to estimate relative social status correctly. In some way, they must have been able to detect the strategy being used by each speaker to convey general politeness to a superior, and to adapt their perception of the speakers' use of vocal cues accordingly. This adaptation may be based on the range of the speakers' voice quality, for example the range of spectral tilt used by a given speaker. However, listeners' adaptation of their politeness perception strategy did not always match the change of the politeness production strategy by speakers. One of the reasons could be that the listeners participating in perception tests did not have familiarity with the speakers' habitual voice usage. Therefore, there was evident consistency between their perceptual judgements and particular acoustic cues. This result suggests that the default acoustic cue to be used generally is the $F3$ region signal such as spectral tilt, when listeners do not have the information about speakers. However, this result might be different if listeners have familiarity with the speaker's voice, and knew the speaker's strategic tendencies in expressing politeness to a superior with whom they are familiar. In this study, we could not necessarily determine whether listeners adapt their politeness judgements strategy according to the speakers' trends in showing politeness.

5.4 Suggestions for future work

This research has shown that in the situation of the Map Task recordings, speakers used *Keigo* according to the relative status of addressee. This supports the suggestion by Jorden (1962) and Inoue (1989) that the use of *teineigo* depends on the required formality rather than the deference or positive politeness to be expressed. This is in opposition to the view of the traditional researchers of the Japanese language who believe that *Keigo* is used to show respect or deference as the word "*Keigo*" literally means. In reality, what happens is that if *Keigo* is not used properly in utterances, the utterances are simply culturally inappropriate, and therefore the speaker is regarded as *im*polite, or

not having proficiency in speaking Japanese (Usami 2002). This in turn disqualifies the speaker as a member of the Japanese society (an MP in the Politeness theory of Brown and Levinson, 1989). An interesting fact is that vocal cues also seem to be used actively for detecting the "*im*politeness". In this study, participants tended to rate tokens as less polite after listening to them than when they read the same tokens in scripts. Not only the inappropriate use of *Keigo* as suggested by Usami, but also the subtle changes of voice, such as a slight spectral tilt change which exceeds the normal range of a speaker, tends to be rated as less polite.

It is possible that overall, people are sensitive to non-polite/*im*polite expressions in written form and speech both, so that they hardly miss any cues which may signal *im*politeness. Therefore, lack of appropriateness of *Keigo* use and failure to control in vocal paralinguistics may both give a bad impression to addressees. If we are developing speech applications such as automatic response telephony system, researchers and developers should pay special attention to this point. So far, the synthesised speech has satisfied lexical appropriateness of politeness, but more importantly, synthesised speech should at least not sound *im*polite, which can be sensed by users quite easily. Therefore, delicate tuning of voice quality is needed to convey politeness properly.

In this study, a variety of vocal changes around the $F3$ region, especially spectral tilt and perturbation of amplitude, were observed according to the change of relative social status of the addressee. However, this variety suggests the possible interference from other types of voice quality characterised by perturbation of F0 and amplitude, and signal to noise ratio. Most of the amplitude and frequency perturbation variation in the $F3$ region came from the irregularity of vocal fold vibration, which is worth paying attention to. Kreiman et al. (2000) mentioned the correlation between shimmer, jitter and hoarseness. In this study, there was no way to confirm this correlation because a larger number of speakers were not available. But in future studies, it will be necessary to consider a number of speakers as well as to consider the involvement of other types of voice quality in conveying politeness.

In the field of speech synthesis, research has focused on the improvement of suprasegmentals such as pitch movement and speech rate, the improved clearness of segmental features in the case of formant synthesis, and the optimisation of segmental unit selection

and join cost in concatenated speech. All these improvements are directed at increasing the naturalness of synthesised speech. However, in order to make speech application users more comfortable when they communicate via synthesised speech, the improved control of paralinguistically meaningful voice quality is also necessary. For example, as listeners' reaction in this study shows, we can possibly infer that the use of extremely steep or shallow spectral tilt should be avoided so as not to give undesirable *im*polite impression. This observation should be taken into consideration in speech synthesis. Also, in the research and development of speech applications, many areas, such as sociolinguistics, pragmatics, and psychology, which had been neglected until recently, should be considered seriously. We observed in this study that even when the speaker's intention is experimentally limited to giving directions to the addressee, the relative social status of the addressee changes the voice quality in ways that are perceptible. If the addressee's social status changes affects voice quality even in these limited contexts, then the situation changes in our real life and the possible strategies based on the politeness theory may be expected to have more salient effects, and should be considered carefully among the social/cultural aspects that must be taken into account in developing synthesised speech. This consideration will include the whole complex of politeness strategies discussed in this thesis. The development of speech applications that are able to adapt to such aspects will improve the performance desired by potential users, such as "polite" responses of automatic speech synthesis/recognition.

This study contributes to information about the association between vocal politeness and some quantitative factors in the higher frequency region, which has been not considered in previous studies. Also, this study shows that it is possible to collect semi-natural spontaneous speech with a certain amount of situational control if needed. As mentioned in section 5.2, it means that this data collection method will be available, not only for the direct measurement of the waveform and spectrum, but also for the source-filter decomposition method which requires high quality recordings.

Hopefully, this study will trigger further studies, which will contribute to the improvement of speech analysis and speech synthesis, including this experimental design with the theoretical approach of pragmatics and sociolinguistics.

Appendix A

Analysis of lexicon (*Keigo* usage)

If the hypothesis suggested by Japanese scholars is true, which is that Japanese show their "politeness" using *Keigo* as a strong cue, *sonkeigo* (respect form) *kenzyoogo* (self humbling form) and *teineigo* (so called "polite form" for showing formality) are likely to be found in their utterances, reflecting their relative social status. On the other hand, if these forms were not found in their utterances, their strategy to show respect might have changed. To examine this aspect, a frequency analysis of each *Keigo* form was conducted.

Materials:

All the utterances of five target speakers were examined, when they participated in the map task dialogues as Instruction Giver. As the target teineigo, the following suffixes were counted.

1. "*-masu*" (an indicative suffix for verbs),
2. "*-desu*" (an indicative suffix for non-verbs such as nouns, adjectives, or adverbs)
3. "*-kudasai*" (a suffix for commands)

Considering turn-taking intervals, each *teineigo* frequency was computed by dividing the total number of mora for the whole dialogue by the total number of *teineigo* suffixes (total morae/total *teineigo* suffixes).

Apart from *teineigo* which is supposed to represents formality, *sonkeigo* and *kenzyoogo*

were also counted separately, when they appeared, because of their rare occurrence.

Results:

Table A.1: Summary of Keigo Frequency

Speaker	Addressee	Number of Mora	desu (A)	masu (B)	kudasai (C)	teineigo (A+B+C)	sonkei+ kenzyoo	teineigo (/mora)
MI	superior(L-H)	4051	61	55	1	117	0	0.0289
	inferior(H-L)	4876	60	53	1	114	0	0.0234
TU	superior(L-H)	3769	53	57	7	117	3	0.0310
	inferior(H-L)	4385	8	9	1	18	0	0.0041
KO	superior(L-H)	3328	23	25	0	48	0	0.0144
	inferior(H-L)	2946	3	7	1	11	0	0.0037
TM	superior(L-H)	6211	143	56	16	215	6	0.0346
	inferior(H-L)	5429	70	48	12	130	1	0.0239
YY	superior(L-H)	6990	114	58	29	201	4	0.0288
	inferior(H-L)	4737	22	5	15	42	0	0.0089

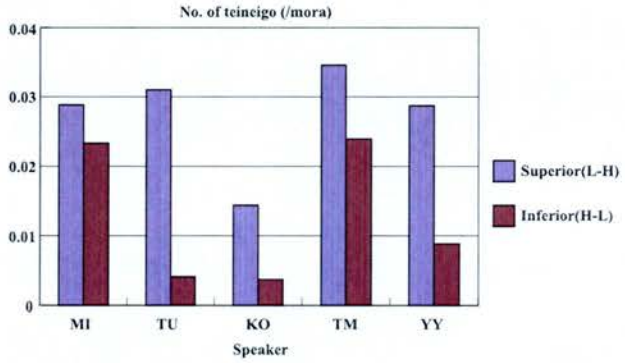


Figure A.1: The effect of Relative Status of addressee and Frequency of Teineigo (/mora)

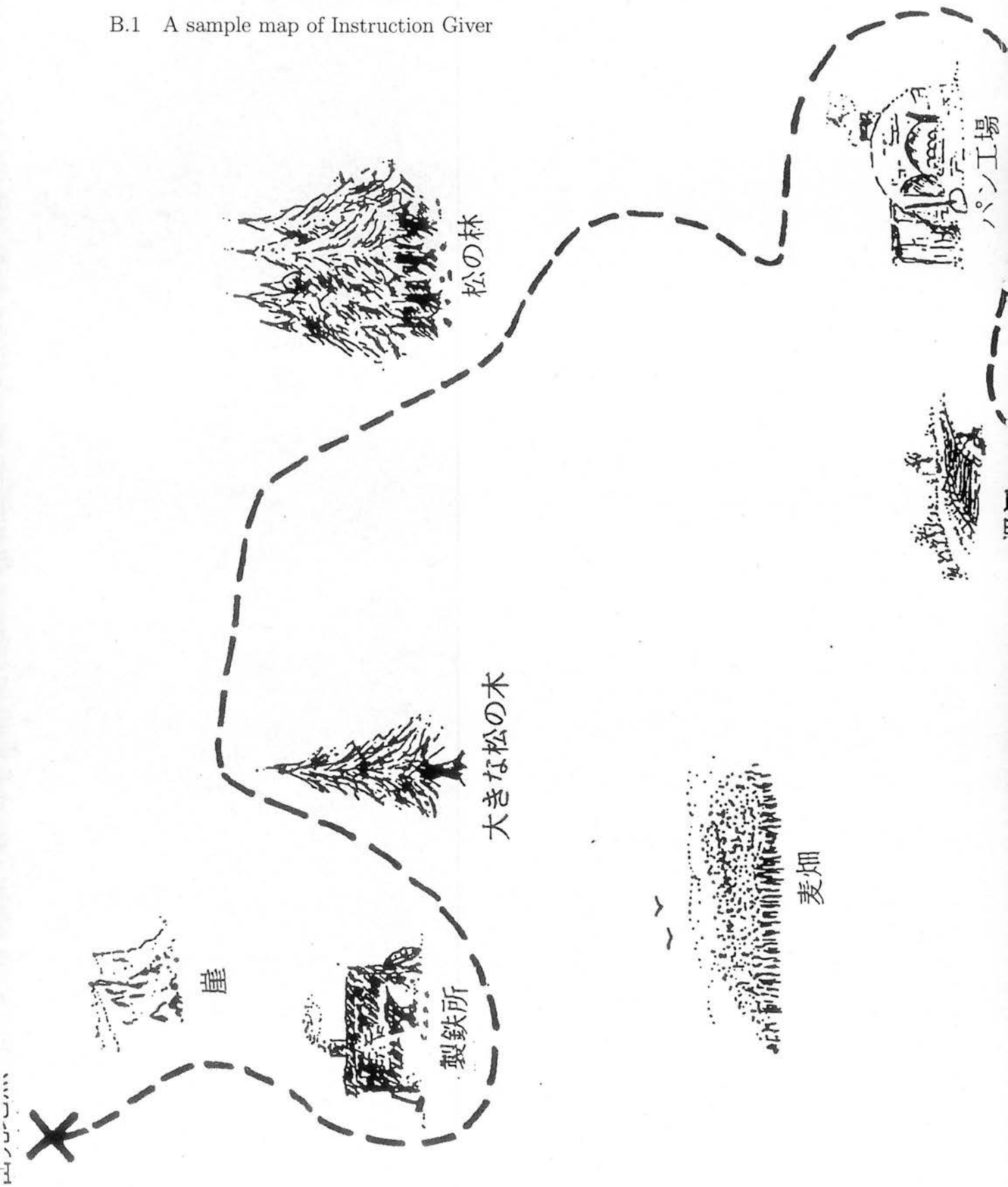
From this result, we can see that very few usages of *sonkeigo* and *kenzyoogo* were found. Some speakers used the presence/absence of *teineigo* to their senior only, but others used *teineigo* to their junior as well. However, from this result, it is difficult to determine whether *teineigo* is used for showing either formality or unfamiliarity. One possible interpretation is as follows. In the case of some speakers who used *teineigo* to their senior only, they kept the rule that *teineigo* should appear all the time when addressing to their superior whereas they can omit the use of *teineigo* to their inferior. In the case of other speakers, who used *teineigo* all the time, they associated *teineigo* with either formality or familiarity, or both of these. In the case of the latter set of speakers, they do not rely on *Keigo* for showing deference, therefore we may consider another channel such as vocal paralinguistics.

Appendix B

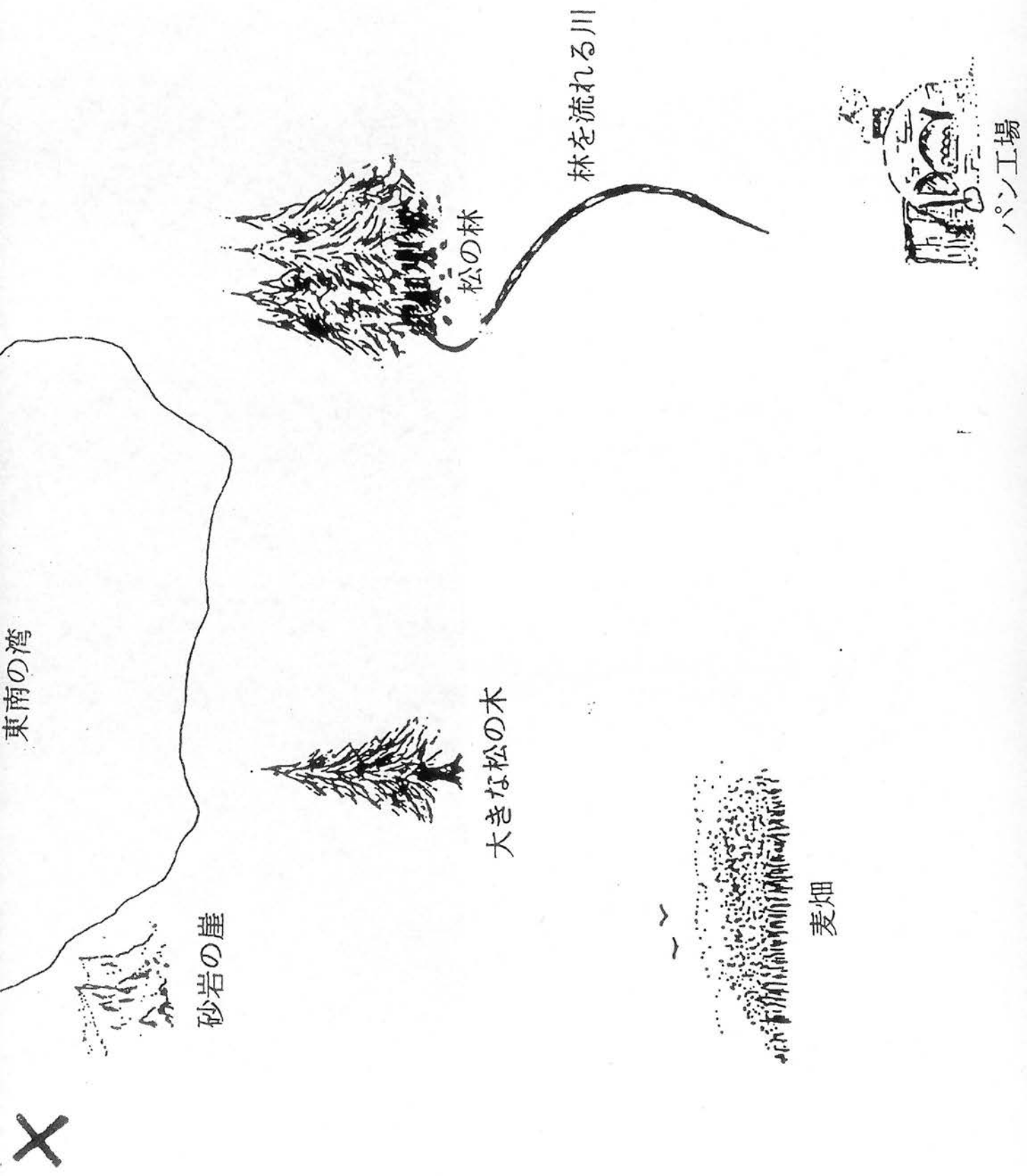
Sample maps used in the Map Task recordings

As shown in Figure 4.2, this study employed the Map Task dialogue recordings. The following two pages include a sample set of maps in the actual size (=A3), one of Instruction Giver with a route, and the other of Instruction Follower without a route.

B.1 A sample map of Instruction Giver



B.2 A sample map of Instruction Follower



References

Addington, D. (1968), 'The relationship of selected vocal characteristics to personality perception', *Speech Monographs* **35**, 492–503.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. & Weinert, R. (1991), 'The HCRC Map Task Corpus', *Language and Speech* **34**, 351–366.

Antoñanzas Barroso, N., Gerratt, B. & Kreiman, J. (2005), 'Synthesizer software for modeling voice quality', *Journal of Acoustical Society of America* **117**, 2544.

Aono, M., Ichikawa, A., Koiso, H., Satoh, S., Naka, M., Tutiya, S., Yagi, K., Watanabe, N., Ishizaki, M., Okada, M., Suzuki, H., Nakano, Y. & Nonaka, K. (1994), 'The japanesse map task corpus: An interim report (in Japanese). *spoken language understanding and discourse processing, japanese society for artificial intelligence.*', *SIG-SLUD-9402* pp. 25–30.

Aronovitch, C. (1976), 'The voice of personality: Stereotyped judgments and their relation to voice qality and sex of speaker', *Journal of Social Psychology* **99**, 207–220.

Bard, E., Robertson, D. & Sorace, A. (1996), 'Magnitude estimation of linguistic acceptability', *Language* **72**, 32–68.

Bezooijen, R. (1995), 'Sociocultural aspects of pitch differences between Japanese and Dutch women', *Language and Speech* **38**(3), 253–265.

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

Buder, E. (2000), Acoustic analysis of voice quality: A tabulation of algorithms 1902-1990, *in* R. Kent & M. Ball, eds, 'Voice Quality Measurement', Singular, San Diego, California.

Campbell, W. (2000), Databases of emotional speech, *in* 'ESCA Workshop on Speech and Emotion', Belfast, pp. 34–37.

Carlson, R., Granstrom, B. & Karlsson, I. (1991), 'Experiments with voice modelling in speech synthesis', *Speech Communication* **10**, 481–489.

Catford, J. (2001), *A Practical Introduction to Phonetics*, second edn, Oxford University Press.

Cranen, B. & Boves, L. (1985), 'Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production', *Journal of Acoustical Society of America* **77**(4), 1543–1551.

Cranen, B. & Boves, L. (1988), 'On the measurements of glottal flow', *Journal of Acoustical Society of America* **84**, 888–900.

de Krom, G. (1993), 'A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals', *Journal of Speech and Hearing Research* **36**, 254–266.

Eckert, P. & McConnell-Ginet, S. (2003), *Language and Gender*, Cambridge University Press, Cambridge.

Epstein, M. (2002), Voice Quality and Prosody in English, PhD thesis, University of California, Los Angeles, Los Angeles, California.

Fant, G., Liljencrants, J. & Lin, Q. (1985), A four-parameter model of glottal flow, Quarterly Progress and Status Report 4, KTH, Speech Transmission Laboratory.

Fant, G. & Lin, Q. (1988), Frequency domain interpretation and derivation of glottal flow parameters, Quarterly Progress and Status Report 2-3, KTH, Speech Transmission Laboratory.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I. & Sundberg, J. (1986), 'Breathiness and insufficient vocal fold closure', *Journal of Phonetics* **14**, 549–553.

Fujisaki, H. (1997), Prosody, models, and spontaneous speech, *in* Y. Sagisaka, N. Campbell & N. Higuchi, eds, 'Computing Prosody: Computational Models for Processing Spontaneous Speech', Springer-Verlag, New York, pp. 27–42.

Gobl, C. (1989), A preliminary study of acoustic voice quality correlates, Quarterly Progress and Status Report 4, KTH, Speech Transmission Laboratory.

Gobl, C. & Ní Chasaide, A. (1988), The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study, Quarterly Progress and Status Report 2-3, KTH, Speech Transmission Laboratory.

Gobl, C. & Ní Chasaide, A. (1992), 'Acoustic characteristics of voice quality', *Speech Communication* **11**, 481–490.

Gobl, C. & Ní Chasaide, A. (1999), Perceptual correlates of source parameters in breathy voice, *in* 'Proc. ICPhS', San Francisco, pp. 2437–2440.

Hammarberg, B., Fritzell, B., Gauffin, J. & Sundberg, J. (1986), 'Acoustic and perceptual analysis of vocal dysfunction', *Journal of Phonetics* **14**, 533–547.

Hanson, H. (1995), Glottal characteristics of female speakers, PhD thesis, Harvard University, Cambridge, MA.

Hanson, H. (1997), 'Glottal characteristics of female speakers: Acoustic correlates', *Journal of Acoustical Society of America* **101**(1), 466–481.

Hanson, H. & Chuang, E. (1999), 'Glottal characteristics of male speakers: Acoustic correlates and comparison with female data', *Journal of Acoustical Society of America* **106**(2), 1064–1077.

Hanson, H., Stevens, K., Kuo, H., Chen, M. & Slifka, J. (2001), 'Towards method of phonation', *Journal of Phonetics* **29**, 451–480.

Hertegård, S., Gauffin, J. & Karlsson, I. (1992), 'Physiological correlates of the inverse filtered flow waveform', *Journal of Voice* **6**(3), 224–234.

Hirose, K., Kawanami, H. & Ihara (1997), Analysis of intonation in emotional speech, *in* 'ESCA Workshop on Intonation: Theory, Models and Applications', Athens Greece, pp. 185–188.

Holmberg, E., Hillman, R., Perkell, J. & Gress, C. (1994*a*), 'Individual variation in measures of voice', *Phonetica* **51**, 30–37.

Holmberg, E., Hillman, R., Perkell, J. & Gress, C. (1994*b*), 'Relationship between intra-speaker variation in aerodynamic measures of voice reproduction and variaton in SPL across repeeated recordings', *Journal of Speech and Hearing Research* **37**, 484–495.

Holmberg, E., Hillman, R., Perkell, J., Guiod, P. & Goldman, S. (1995), 'Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice', *Journal of Speech and Hearing Research* **38**, 1212–1223.

House, A. & Stevens, K. (1958), 'Estimation of formant bandwidths from measurements of transient response of the vocal tract', *Journal of Speech and Hearing Research* **1**, 309–315.

Ide, S. (1989), 'Formal forms and discernment: two neglected aspects of linguistic politeness', *Multilingua* **8**(2/3), 223–248.

Inoue, F. (1989), *Kotobadukai Shin Fuukei(Keigo to Hyogen) (New Landscape of Language Use (Honorifics and Dialect))*, Akiyama Shoten, Tokyo, Japan.

Inoue, F. (1998), *Nihongo Watching (Japanese Language Watching)*, Iwanami Shoten, Tokyo, Japan.

Ito, M. (2002), Japanese politeness and suprasegmentals: A study based on natural speech materials, *in* B. Bel & I. Marlien, eds, 'Proc. of the Speech Prosody 2002 conference', Aix-en-Provence, pp. 415–418.

Jorden, E. (1962), *Beginning Japanese*, Vol. 1, Yale University Press.

Karlsson, I. (1991), 'Female voices in speech synthesis', *Journal of Phonetics* **19**, 111–120.

Kent, R. (1997), *The Speech Sciences*, Singular, San Diego, California.

Klatt, D. & Klatt, L. (1990), 'Analysis, synthesis, and perception of voice quality variations among female and male talkers', *Journal of Acoustical Society of America* **87**(2), 820–857.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. & Den, Y. (1998), 'An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs', *Language and Speech* **41**(3-4), 295–321.

Kreiman, J. & Gerratt, B. (1998), 'Validity of rating scale measures of voice quality', *Journal of Acoustical Society of America* **104**(3), 1598–1608.

Kreiman, J. & Gerratt, B. (2000*a*), Measuring vocal quality, *in* R. Kent & M. Ball, eds, 'Voice Quality Measurement', Singular, San Diego, California, pp. 74–101.

Kreiman, J. & Gerratt, B. (2000*b*), 'Sources of listener disagreement in voice quality assessment', *Journal of Acoustical Society of America* **108**(4), 1867–1875.

Ladefoged, P. (2001), *A Course in Phonetics*, fourth edn, Harcourt College Publishers.

Ladefoged, P. & Antoñanzas Barroso, N. (1985), Computer measures of breathy voice quality, Working Papers in Phonetics 61, University of California at Los Angeles.

Laver, J. (1980), *The phonetic description of voice quality*, Cambridge University Press, Cambridge.

Laver, J. (1991), *The Gift of Speech*, Edinburgh University Press, Edinburgh, Scotland.

Laver, J. (1994), *Principles of Phonetics*, Cambridge University Press, Cambridge.

Laver, J. (2000), Phonetic evaluation of voice quality, *in* R. Kent & M. Ball, eds, 'Voice Quality Measurement', Singular, San Diego, California.

Machida, K. (1999), *Tsurutsuru No Tsubo*, Kodansha, Tokyo, Japan.

Matsumoto, Y. (1988), 'Reexamination of the universality of face: Politeness phenomena in Japanese', *Journal of Pragmatics* **12**, 403–426.

Moore, W. (1939), 'Personality traits and voice quality deficiencies', *Journal of Speech and Hearing Disorders* **4**, 33–36.

Murphy, P. (1999), 'Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis', *Journal of Acoustical Society of America* **105**, 2866–2881.

Murphy, P. (2000), 'Spectral characterization of jitter, shimmer, and additive noise in synthetically generated voice signals', *Journal of Acoustical Society of America* **107**, 978–988.

Ní Chasaide, A. & Gobl, C. (1993), 'Contextual variation of the vowel voice source as a function of adjacent consonants', *Language and Speech* **36**, 303–330.

Ní Chasaide, A. & Gobl, C. (1997), Voice source variation, *in* W. Hardcastle & J. Laver, eds, 'The Handbook of Phonetic Sciences', Blackwell, pp. 427–461.

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

Ogino, T. & Hong, M. (1992), Nihongo onsei no teineisa ni kansuru kenkyuu (a study on politeness in Japanese speech), *in* M. Kunihiro, ed., 'Nihongo intonation no jittai to bunseki (The State-of-the-art and Analysis of Japanese Intonation)', Monbushou (The ministry of Education), pp. 215–258.

Ohala, J. (1984), 'An ethlogical percpective on common cross-language utilization of f_0 of voice', *Phonetica* **41**, 1–16.

Ohala, J. (1996), Ethlogical theory and the expression of emotion in the voice, *in* 'Proc. ICSLP', Vol. 3, Philadelphia, pp. 1812–1815.

Ohara, Y. (1999), Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English, *in* U. Pasero & F. Braun, eds, 'Wahrnehmung und Herstellung von Geschlecht', Westdeutscher Verlag, pp. 105–116.

Ohara, Y. (2000), A critical discourse analysis: Ideology of language and gender in Japanese, PhD thesis, University of Hawai'i at Manoa.

Ohara, Y. (2001), Finding one's voice in Japanese: A sutdy of the pitch levels of L2 users, *in* A. Pavlenko, A. Blackledge, I. Piller & M. Teutsch-Dwyer, eds, 'Multilingualism, Second Language Learning, and Gender', Mouton de Gruyter, pp. 223–248.

Qi, Y. & Hillman, R. (1997), 'Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals', *Journal of Acoustical Society of America* **102**, 537–543.

Shibatani, M. (1990), *The languages of Japan*, Cambridge University Press, Cambridge.

Shrivastav, R. (2001), Perceptual structure of breathy voice quality and auditory modeling of its acoustic cues, PhD thesis, Indiana University, Bloomington, Indiana.

Shrivastav, R. (2003), 'The use of an auditory model in predicting perceptual ratings of breathy voice quality', *Journal of Voice* **17**(4), 502–512.

Södersten, M. & Lindestad, P. (1990), 'Glottal closure and perceived breathiness during phonation in normally speaking subjects', *Journal of Speech and Hearing Research* **33**, 601–611.

Södersten, M., Lindestad, P. & Hammarberg, B. (1991), Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers, *in* J. Gauffin & B. Hammarberg, eds, 'Vocal Fold Physiology : Acoustic, Perceptual, and Physiological Aspects of Voice Mechanism', Singular, San Diego, California, pp. 217–224.

Stevens, S. (1969*a*), 'The direct estimation of sensory magnitudes – loudness', *American Journal of Psychology* **69**, 1–25.

Stevens, S. (1969*b*), 'On predicting exponents for cross-modality matches', *Perception and Psychophysics* **6**, 251–256.

Titze, I. (1994), *Principles of Voice Production*, Prentice-Hall, New Jersey.

Usami, M. (1999), Discourse Politeness in Japanese Conversation, PhD thesis, University of Harvard.

Usami, M. (2002), 'Poraitonesu riron no tenkai (the development of politeness theory)', *Gekkan Gengo* .