**University of Alberta**

# Recipient Response Behaviour during Japanese Storytelling: A Combined Quantitative/Multimodal Approach

by

Neill Lindsey Walker

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts
in
Japanese Language and Linguistics

Department of East Asian Studies

Fall 2010
Edmonton, Alberta

## Examining Committee

Tsuyoshi Ono, East Asian Studies

Anne Commons, East Asian Studies

Geneviève Maheux-Pelletier, Modern Languages and Cultural Studies

***For Sho***

# Acknowledgement

I would like to thank my supervisor, Professor Tsuyoshi Ono for his valuable knowledge, input, and coffee as well as for the generous use of his corpus data for this and a variety of other projects. I would also like to thank Professor Anne Commons for her support and friendship, and Professor Geneviève Maheux-Pelletier for asking the hard questions that always seem to lead me to find better answers.

This thesis would not have been possible without the support of the wonderful team of faculty and staff in the Department of East Asian Studies, and the volunteers who participated in this study. Thank you all.

## Abstract

This study explores the role of speaker and listener gaze in the production of recipient responses, often called backchannels or, in Japanese, *aizuchi*. Using elicited narrative audio/video data, speaker gaze and recipient response behaviours were first analyzed quantitatively. The results showed that majority of recipient responses are made while the speaker is gazing at the recipient. Next, a qualitative multimodal analysis was performed on a specific type of recipient response that occurred both during and without speaker gaze. The results showed that recipients make good use of the state of the speaker's gaze to regulate the speaker's talk and negotiate for a pause, a repair, or a turn at talk. These findings suggest that what are currently known as backchannels are only a small part of a much larger sequential multimodal system that is inseparable from the ongoing talk.

# Table of Contents

List of Tables ........ 9

List of Figures ........ 9

1 Introduction ........ 10

1.1 Introduction to ‘Backchannels’ ........ 10

1.2 Seminal backchannel studies ........ 13

1.3 Issues in backchannel studies ........ 14

1.4 Backchannel studies in Japanese and *aizuchi* ........ 15

1.5 Other Approaches ........ 24

1.6 The Current Study ........ 27

2 Data ........ 33

2.1 The Pear Film ........ 33

2.2 Participants ........ 34

2.3 Methodology ........ 35

2.4 On the use of elicited data ........ 37

2.5 Coding and Analysis ........ 38

3 Quantitative Analysis ........ 42

3.1 Recipient responses and speaker head nods ........ 42

3.2 Results for the Quantitative Analysis of Gaze ........ 51

3.3 Summary of Quantitative Results ........ 56

4 Qualitative Analysis ........ 57

4.1 Initial Observations on Gaze and Interaction ........ 61

4.2 Speaker Gaze and Recipient Responses ........ 63

4.2.1 YU-AZ Selection ........ 63

4.3 Recipient Responses and Speaker Pause ........ 68

4.3.1 A-N Transcript 1 ........ 69

4.3.2 A-N Transcript 2 ........ 71

4.3.3 A-N Transcript 1 + gaze information ........ 71

4.4 Recipient Responses and Speaker Repair (Increment) ........ 74

4.4.1 A-N Transcript ........ 75

4.4.2 A-N Video stills ........ 76

4.5 Recipient Gaze Change as Turn-taking Device ........ 81

4.5.1 A-YM Transcript ........ 82

4.6 A previously overlooked response: “Head Twist” ........ 87

4.6.1 YU-AZ Video Stills ........ 88

4.7 Summary of Qualitative Results ........ 91

5 Conclusions ........ 93

5.1 Backchannels? *Aizuchi*? ........ 95

5.2 Gaze ........ 96

5.3 Baseline Behaviour and Deviations from the Baseline ........ 97

5.4 Issues: video & technology ........ 99

5.5 Issues: transcription ........ 99

5.6 Further implications ........ 101

6 Works Cited ........ 103

7 Appendices ........ 108

7.1 Appendix 1 – Transcription symbols ........ 108

7.2 Appendix 2 – Equipment ........ 108

## List of Tables

Table 1: Comparison of this Thesis with Previous Studies ........ 28

Table 2: Participants ........ 35

Table 3: Quantitative Results of Recipient Responses by Type ........ 44

Table 4: Quantitative Results for Speaker Gaze ........ 53

Table 5: Recipient Manipulation of Speaker Gaze ........ 60

## List of Figures

Figure 1: Utilizing visually available resources ........ 31

Figure 2: Gaze unit sequence ........ 40

# 1 Introduction

In this thesis I examine recipient response behaviour, often referred to as 'backchannels' or *aizuchi*, in Japanese talk. Utilizing a small video corpus of elicited narrative conversations in which one participant was shown a short film and asked to describe it to a second participant, I first take a quantitative approach to examine recipient production of verbal and nonverbal responses and their relation to speaker gaze. Then I perform a close qualitative analysis of some response behaviours that have generally been overlooked. The qualitative analysis takes a multimodal approach, factoring in all the actions, verbal and non-, that speakers and recipients make, as well as their interactions with the environment.

I will look at some specific examples in detail to show that recipient responses are in fact very complex coordinated behaviours, and that they are closely tied to the state of speaker gaze. I will argue that investigators looking at backchannels, especially in a Japanese context, need to broaden the scope of behaviours that are analyzed, to consider the multi-modal nature of backchannels, and to put a greater reliance on video data. The results of this study will, I hope, make others question the validity of the terms '*aizuchi*' and 'backchannel,' and underline the importance of including nonverbal aspects of conversation in language studies.

## 1.1 Introduction to 'Backchannels'

Backchannels are usually considered to be short utterances that are made during a speaker's turn at talk, but that do not result in a turn change. Typical English examples are “uh-huh” and “yeah.” In Japanese, *hai* and *ee* “yes” and *un* “uh-huh” are very common.[1] Below is an example of a backchannel in English from Ynvge (1970:574):

**Excerpt 1**

```
   1  A:There’s a fellow here, Bob Fabry,
→  2  B:Uh-huh
   3  A:who’s been writing his dissertation,
```

And in Japanese (Hatasa, Hatasa, and Makino, 2011:118):

**Excerpt 2**

```
  1  A:  kinoo     eiga    ga    atte,
          Yesterday movie  SUB  exist-and
→ 2  B:  ee,
          yes
  3  A:  omoshiroi  tte    yuu kara itte   mita n  desu  ga,
          Interesting QUOT   say so   go-and  see     COP   but
```

Translation (ibid):
1 A: There was a movie yesterday and,
2 B: Yes.
3 A: I went because I heard it would be interesting,

From the examples above, backchannels may seem to be a category that is quite simple to define. The examples are misleading, however, because of the wide variety of responses that occur in the same environment, ranging from single words (or sometimes even sounds) to grammatically complete sentences. Laughter, too, could be considered a backchannel. In addition, Yngve (1970) briefly describes a variety of nonverbal behaviours, including gesture, that could fall within the category of backchannels, too. Several investigators looking at Japanese, most notably Maynard (1989), include head nodding as a nonverbal example of backchanneling in that language. To further complicate matters, not all backchannels, verbal or nonverbal, are produced as neatly as the above examples might imply; they can and do overlap the speaker’s talk or are produced iteratively and seemingly randomly in terms of form (uh huh, yes, ok, sure, etc).

[1] A more thorough look at various *aizuchi* will be carried out in the section on *aizuchi*, below.

The topic of ‘backchannels’ and all the behaviours that could possibly fall under this term, both verbal and nonverbal, seem to be problematic in that they fall somewhere between the realms of gesture and talk, uncomfortably outside the interest areas of both gesture researchers and linguists. Schegloff (1981) refers to backchannels as “detritus” (74) that language investigators, until that point, had ignored in favour of “cognitive structure and quasi-syntactic composition of discourse” (ibid), meaning they fell outside the range of behaviours that researchers, and their resulting methodologies, found to be of interest. The term “backchannel” can perhaps be considered waste-bin taxonomy for a wide range of utterances and behaviours that initially were not of interest to linguists.  This has resulted in a variety of studies that look at backchannel behaviour, but as a group they lack coherence due to differing methodologies, type of data analyzed, and disagreement over the definition of the term ‘backchannel.’ This has resulted in an overabundance of methods, explanations, and jargon. Overwhelmingly the data used has been limited to audio recordings (or transcripts of audio recordings), which may be limiting our understanding of the complex nonverbal elements of talk, as demonstrated by papers such as Goodwin (1981) or Ford, Fox, and Thompson (2002). In the following subsections I will introduce some previous backchannel studies, starting with Ynvge (1970), widely considered to be the first paper specifically on backchannels. I will then discuss in detail some of the issues that exist within backchannel studies. Following that I will give an overview of some major Japanese-specific backchannel studies (often called *aizuchi* in the Japanese case) and the additional issues that have resulted from them.

## 1.2 Seminal backchannel studies

Yngve (1970) is credited with coining the term “backchannel,” and most subsequent papers on the subject have cited Yngve for the purpose of either elaborating on or refuting his observations and definitions, which were actually based on his “preliminary analysis” (573) of data. Yngve’s data collection method was very advanced for the time, utilizing both audio and film, as was the scope of his analysis, which included nearly all elements of conversation, including verbal utterances, nonverbal gesture, body movements, and gaze, though it did not fall into an accepted methodological framework.[2] However complex and innovative Yngve’s study was, however, his definition of what exactly constitutes a “backchannel” and the environment it occurs in is unclear, which has had great consequences for later studies.

Schegloff (1981) responds to Yngve (1970)’s paper and the resultant approach to backchannels from a talk-in-interaction framework. He criticises Yngve (and other language investigators in general) for treating backchannels as if they were a special class of behaviours that are separate from the rest of what goes on in language. He rightly notes that it is the environment recipient responses occur in that makes them relevant and meaningful to the talk at hand, and that there is no acceptable way to demarcate them (lexically, syntactically, etc.) from what else goes on in conversation. Schegloff calls these responses “continuers” in that they are used to pass up a turn at talk; in other words, to allow the speaker to extend the unit of talk[3], and includes them in “the organizational domain of repair” (91). The study was based on audio recordings

[2] It is likely that no such framework existed at the time. Yngve’s own framework was called “state of mind” and is a turn-based approach similar in some ways to CA, which coincidentally emerged at roughly the same time.

[3] A unit of talk is whatever the participants collaboratively create; a turn, a sentence, a word, etc.

of telephone conversations, and specifically utterances of “uh-huh[4],” thus representing a narrower data set than Ynvge (1970).

## 1.3 Issues in backchannel studies

Although Yngve’s “backchannel” has become the more popular term, it is the fundamental differences between Yngve (1970) and Schegloff’s (1981) analyses that continue to be debated in nearly all subsequent papers on the topic. The underlying methodologies employed by these two investigators lies at the root of the disagreement, resulting in several issues which are addressed by Gardner (2001). One of these issues, mentioned briefly above, is an overabundance of jargon stemming from attempts to functionally describe backchannels. Gardner, in his literature review, lists twelve new terms for backchannels that are meant to refine or refute the generic terms “backchannel” and “continuer” based on functional analyses. Unfortunately, the studies Gardner lists display a wide degree of variation in terms of data (audio vs. video, natural conversation vs. elicited, natural vs. experimental settings, etc) and methodology. Gardner is critical of previous studies:

> “… it will be seen that there is a lack of consensus but also it is suggested that this is, at least in part, because these diverse and highly variable utterance types are frequently treated as a homologous and undifferentiated group, lumped together through premature coding, with accompanying claims that they display, for example, attention, agreement or understanding” (2001:17).

The first criticism deals with the fundamental issue in backchannel studies: what constitutes a backchannel has never been clearly defined, leading to a haphazard

[4] Schegloff (1981) also refers to “other things” in addition to “uh-huh” but what exactly constitutes “other things” is left undefined.

grouping of behaviours under an umbrella term. The second is that of transcription. What Gardner calls "premature coding" is the tendency in backchannel studies to adopt the coding methods of previous studies without regard to differences inherent in the data, which range from type of media (audio, video) genre (interview, elicited narrative, natural conversation), participants (age, gender), location (school, work, café), or even language. Because of this, backchannel studies have assumed, at least until recently, that English "uh-huh" is analogous to "un" in Japanese, or that a head nod carries the same meaning/function across languages.[5]

## 1.4 Backchannel studies in Japanese and *aizuchi*

"*Aizuchi*" is a word that comes up often in Japanese backchannel studies. Depending on the author, it may be used to distinguish Japanese backchannels from those other languages, describe backchannels in the context of the Japanese language, or be used interchangeably with the word 'backchannel.' The original or archaic meaning of *aizuchi* refers to the motions made by two blacksmiths facing each other and hammering metal in turn. How it developed into a language-related term is not clear, but the action it implies, a rhythmic bowing or nodding in the course of hammering steel, is salient in talk between Japanese native speakers. Like the term 'backchannel,' '*aizuchi'* is, unfortunately, undefined (Yamada 2004), and denotes a variety of different forms and behaviours depending on the context it is used in. The one thing most reference books seem to agree on is that *aizuchi* are 'responses[6] of

[5] But see McClave, Kim, Tamer, and Mileff (2007) for a cross-cultural analysis of speaker head nods.

[6] Increasingly, however, investigators are looking at *speaker* behaviours that resemble backchannels, particularly but not limited to head nods. Such behaviors are clearly not response behaviours. See, for example, Iwasaki (1997), Hayashi (2002) (though he does not use the terms "backchannels" or "*aizuchi*"), McClave et al (2007), and Aoki (2008).

agreement[7]’, although some investigators claim a wider range of meaning[8]. They are also supposedly a requirement of good conversation in Japanese: “A good listener,” announces a small grammar book for children from the 1950’s, “should interject aizuchi because they are like oil on the axle of a wheel- they make the wheel of conversation go around smoothly. [9]”

*Aizuchi* are described in various ways depending on the source. Older sources, such as the example from Yoshida (1951), above, as well as many textbooks, typically refer to expressions such as:

| | | | |
|---|---|---|---|
| soo desu ka | “is that so?” | soo desu ne | “yes, that’s right” |
| naruhodo desu | “I see” | hontoo desu ka | “really?” |

These expressions have clear semantic meaning and are grammatically complete sentences. They are the kind of ‘textbook *aizuchi*’ that one is likely to learn in a Japanese language class, due to their perceived level of politeness or situational appropriateness. This is not an exhaustive list, of course. Interestingly a couple of them, “*soo desu ka*” and “*hontoo desu ka*”do not necessarily show agreement[10]. In my data set these phrasal *aizuchi* do not appear[11], and they are probably much less common in natural conversation than the following, much shorter utterances that can also be used as *aizuch*[12]*i*:

[7] The *Kojien* (2001) dictionary uses the term *chooshi o awaseru*, roughly “to harmonize” or “be agreeable.”

[8] Ward and Tsukahara (1999) claim that some backchannels in Japanese display boredom and scepticism. Tanaka (2004) mentions ‘negative’ *aizuchi* in her discussion of gender variables (193).

[9] Yoshida (1951:79), quoted by Miller (1983). Conversation with Miller by email, December 11, 2009.

[10] Kita and Ide (2007) say that some *aizuchi* work as ‘assessments’ of a speaker’s talk.

[11] This could be due to the genre of the conversation (elicited narrative), the topic of conversation, the age or relationship of the participants, or some other unknown factor.

[12] When these utterances are produced during another speaker’s turn at talk, they are considered to be *aizuchi*. When they occur as part of a speaker’s own turn at talk, for example

| | | | |
|---|---|---|---|
| hai | “yes” | ee | “yes” |
| un/n/m[13] | “uh huh” | e?/ee? | “what?” |

These expressions are shorter, and appear more similar in meaning to the “uh-huhs” and “yeahs” Schegloff (1981) analyzed. These forms may be more colloquial (except for *hai*, which is considered polite) and very likely more frequent in typical conversation. Although I did not systematically count every utterance for the purpose for comparison, it was obvious that *un* and its variants, *n* and *m* were the most frequently produced *aizuchi* in the data set. “*Hai*” was produced by only one speaker[14]. The list above is far from exhaustive. Martin (2004), lists the above examples along with about forty other words and phrases, and their numerous variations[15] under the term ‘interjections’ that can and do occur in the same speech environment as the *aizuchi* noted above. He does not specifically mention *aizuchi* at all.[16]. Japanese aizuchi/backchannel studies are prolific, though it appears that a great many of them focus on SLA or Japanese foreign language education. This stems from the perception that *aizuchi* is a socially important behaviour in Japanese and is necessary for cross cultural communication (Lebra 1976, Yamada 1992), and has a basis in social cohesiveness (Kita and Ide 2007) and the desire to avoid disagreement or conflict (Yamada 1992) while promoting or maintaining harmony (Kogure 2007). *Aizuchi*, then,

when answering a question or as a filler (or other device) during a turn at talk, they are not considered to be *aizuchi*.

[13] These utterances are usually transcribed as “un,” yet are often pronounced as either “n” (nasal, mouth open) or “m” (mouth completely closed). To my knowledge no one has looked into patterns of use of these three articulations, though Gardner (2001) discusses it in his critique of transcription accuracy. I have attempted to accurately reflect the actual pronunciation in the transcriptions here based on careful observation of participants’ lips.

[14] Probably due to perceived social distance with the speaker. More discussion on this is below.

[15] Many of the expressions listed by Martin have up to four alternate forms based on variations in pronunciation and vowel length. I have chosen only a few of the representative ones here.

[16] Martin (2004) discusses “interjection-type” particles (specifically *ne* and *na*) that are “used to involve both speakers and hearers in what is being said” (914).

have been attributed with a variety of functions that have only tenuous connection to their actual production within a conversation and at times these claims seem ethnocentrically charged[17]. According to Tanaka (2004), there is a danger that such nationalistic bias can be retained in academic work on *aizuchi*, either intentionally or unintentionally. Likewise Maynard (1997) and White (1989) criticize many previous studies for a lack of empirical data, non-standard methodological approaches, and a reliance on anecdotal evidence. This suggests that when empirical methods fail to explain *aizuchi*, there may be a tendency to fall back on hearsay or popular notions of the functions of *aizuchi*, which are motivated by a form of cultural protectionism which strives to prevent empirical analysis[18]. The term *aizuchi* is thus saddled with a set of very problematic issues, from a lack of consensus on definition to a questionable base of evidence, and these are often compounded by the issues also inherent in the term “backchannel.” However, several researchers have made efforts to approach Japanese backchannels/*aizuchi* systematically.

Maynard (1989) may have been the first to look at backchannels (including head nods) systematically in Japanese, based on Yngve’s (1970) definition. Due to the ambiguity present in Yngve’s original description of backchannels and the complexity of backchannel production, Maynard carefully limited the focus of her study to specific environments and behaviours[19]. She found that the majority of Japanese backchannels

[17] Miller (1982) and Tanaka (1999) point out themes of social harmony in national discussions of the Japanese language stemming from *Kokugogaku*, the field of traditional Japanese linguistics, and *Nihonjinron*, “theories of Japaneseness,” and the danger these themes represent to objective analyses of Japanese language and culture.

[18] Lee (2010) discusses the continuing historical struggle between *Kokugogaku* scholars, who strive to protect traditional values and notions of the Japanese language, and Western-style linguists, who strive to approach the same problems empirically.

[19] Backchannels are limited to those produced by the recipient during the speaker’s current turn at talk and those produced after a speaker’s talk and then followed by a pause before the listener takes a turn (161). Nonverbal backchannels are limited to vertical nods or horizontal shakes (161).

are produced at the boundaries of what she calls "pause-bounded phrase units" (PPUs)[20]. Maynard (1990) compared backchannel behaviours in American English and Japanese and found that Japanese make backchannels about twice as frequently as Americans, and that 38.08% of Japanese backchannels occur after speaker head movement occurs, compared to only 7.78% in the English case[21]. Maynard (1997) built on her previous studies, using a similar approach (Contrastive Conversation Analysis) and attempted to control for sociolinguistic variables in the data. Her studies have a high degree of replicability which, unfortunately, many backchannel studies seem to lack.

Maynard's (1989, 1990, 1997) findings on frequency have spawned numerous cross-linguistic comparison studies on backchannel frequency. However it is unclear what cross-cultural comparisons on frequency can tell us, especially since backchannels in Japanese have been shown to occur in different speech environments than, for example, English[22] (Kita and Ide 2007), and some authors doubt an analogous relationship between Japanese *aizuchi* utterances and English "uh-huhs" and "yeahs" (ibid). Very few subsequent Japanese backchannel studies have attempted to build systematically on Maynard's methodology, though her findings are widely quoted. And while Maynard's inclusion of (vertical) head nods in her study introduced, perhaps for the first time since Yngve (1970), a nonverbal component to the investigation of

[20] A PPU is a unit of talk occurring within a single intonation contour, bounded by identifiable pause (Maynard 1989, 163). Maynard demonstrates that Japanese units of talk are produced in regular phrase-level chunks, in contrast to the sentential-level chunks associated with English talk. See also Iwasaki (1993). The PPU is similar to the intonation unit (IU) discussed by Chafe (1994).

[21] This observation has only been followed up on recently, by Aoki (2008).

[22] Backchannels occur mid-turn in Japanese, rather than at transition relevance places (TRPs) (Kita and Ide 2007), although it may be that TRPs in Japanese are tied to prosodic units, for example PPUs (Maynard 1989) or the intonation units described by Iwasaki (1993). TRP describes a place at the end of a turn construction unit (TCU), the fundamental unit of conversation in CA, where a turn change may occur (Sacks, Schegloff, and Jefferson 1976).

Japanese backchannels, very few subsequent papers attempt to go beyond recipient head nods[23]. I believe this has resulted in a biased assumption of what behaviours constitute Japanese backchannels; a case of "premature coding" (Gardner 2001) in which the arbitrary selection of one type of behaviour has led to other behaviours being ignored. And there is other bias: *aizuchi* – head nodding and short vocalizations by a recipient during a speaker's ongoing talk, to use one definition – is a topic for study in the first place because such behaviour is presented as being qualitatively different from what others (specifically, English speakers) do[24]. Thus statements such as "Japanese perform *aizuchi* more frequently[25] than speakers of other languages such as English and Mandarin[26]" (Tanaka 2004:137; also Kita and Ide 2007) are curious, since *aizuchi* are often typified as being specifically Japanese behaviour[27]. It is possible (and even likely) that English native speakers are performing different actions, at different times, that are functionally analogous to Japanese backchannels, but they simply fall out of the range of behaviours arbitrarily selected by comparison studies, or, as Kita and Ide say, they may have not yet been described (2007:1252). In the following I will present a selection of papers which represent a variety of approaches to the study of Japanese backchannels or *aizuchi*.

[23] Notable exceptions are Szatrowski (2003), who also looks at gaze, and Kogure (2007) who includes smiling in her analysis. Other recent studies such as Hayashi (2002) and Iwasaki (2009) include a wide variety of nonverbal behaviors, but do not use the term "backchannel" at all.

[24] As in Lebra (1987). Sugito (1987) discusses the "foreigner's view" in his introduction.

[25] Maynard (1997) notes that a large number of papers that make cross-linguistic claims about *aizuchi* frequency are not methodologically qualified to make quantitative claims.

[26] Tao and Thompson (1990), the authors of the paper cited in Kita and Ide (2007) and Tanaka (2004) used the term 'backchannels,' not '*aizuchi*,' and they defined backchannels as "any change in speakership, whether in overlap or not" (211). In other words, their data is audio only and their definition of 'backchannel' differs significantly from that of *aizuchi* in many other studies, making direct quantitative comparisons suspect.

[27] For further examples, see Yamada (1992) on frequency and sociocultural meaning, and Miller (1991:99) on differences in production.

An interesting study that takes a different approach to Japanese backchannels is Ward and Tsukahara (2000), who examined the effect of speaker prosody on recipient production of backchannels. Using a corpus of Japanese conversation data in which the participants could not see each other, they found that 110 milliseconds of low pitch is the best predictor of backchannel production by recipients and claim that the semantic content of speaker talk has less correlation to backchannel production than was previously assumed. Syntactically, however, they found low pitch areas to often coincide with phrase endings, grammatical completion, and the use of so-called 'final particles[28].' On the other hand, Koiso et al (1998), using a map task corpus in which some participants could see each other and others could not, argue that syntactic cues are more relevant than any specific incidence of low pitch prosody, based on the appearance of interjection, case, and final particles as well as intonation curves (rather than just pitch) in their analysis. The differences between Ward and Tsukahara (2000) and Koiso et al (1998) could stem from the way they identified pitch or due to differences in the genre of talk they analyzed.

Kita and Ide (2007) argue against the use of terms such as "backchannels" or "continuers" in the Japanese context, preferring what they say is the "theory neutral" term, *aizuchi* (1243)[29]. Finding, for example, that many Japanese backchannels occur away from transition relevance places (TRPs)[30], they dispute the accuracy of Schegloff's

[28] 'Final particles,' include *ne*, *na*, yo, *zo*, *ze*, and others. Unlike the grammatical case particles, they do not have a grammatical function. The name probably stems from their appearance at the end of sentences in written language. More recent linguistic analyses based on natural speech led to new names, such as 'interactional particles' or 'conversational particles.' They tend to appear at the end of phrases or clauses, although they can also appear after bare nouns, and sometimes on their own.

[29] But see above for evidence that the term *aizuchi* is anything but theory-neutral.

[30] TRP describes a place at the end of a turn construction unit (TCU), the fundamental unit of conversation in CA, where a turn change may occur (Sacks, Schegloff, and Jefferson 1976).

(1982) term, “continuer.”[31] Kita and Ide argue that Japanese backchannels are a kind of “phatic communion” that represent “coordination for the sake of coordination” (1250) within the Japanese cultural construct, very different from the simple context-derived responses that make up ‘backchannels.’ Their claim that *aizuchi* are Japanese-specific tools for consideration and cooperation (1251) place *aizuchi* out of the reach of linguistic analysis and echo the ethnocentric tendencies illustrated by Miller (1982) and Tanaka (2004).

Despite their obvious differences, all the above papers approach backchannels and *aizuchi* in Japanese with the assumption that such behaviour is performed solely by recipients in response to a speaker’s talk. One paper that takes a more interactive approach is Iwasaki (1997). He describes cases where a “loop” sequence of backchannels between speaker and recipient can result in a turn change. The loop sequence occurs when a speaker ends an utterance and the recipient responds by producing a backchannel, which normally prompts a speaker to continue. However in this case, the speaker produces a backchannel response too, which gives the recipient a turn at talk. Iwasaki’s paper is significant also for his criteria of backchannels, which have been adopted by other studies. He divides backchannels into three categories: non-lexical (short sounds with little or no referential meaning, usually continuers); phrasal (one- or two-word stereotypical expressions), and substantive (sentence or series of sentences) (666).

Using Iwasaki’s (1997) criteria, Kogure (2008) looked at nonverbal behaviours such as nodding and smiling which occur during the loop sequence in Japanese. Kogure found that Japanese conversationalists try to avoid silence and provide

[31] Schegloff claimed, however, that it is possible that continuers and other repair-related devices can occur at any point during a turn.

feedback through the use of verbal and nonverbal actions during the loop sequence for the purpose of maintaining “harmony”. This finding conflicts with stereotypical views on silence as an example of harmony in Japanese as claimed by Lebra (1987)[32], and like Kita and Ide (2007) and Yamada (1992), is another paper that stresses “harmony,” “cooperation” and “avoidance of conflict” as a natural function of Japanese *aizuchi*.

Szatrowski (2002) is one of the few *aizuchi* investigators who explore the role of gaze in the production of *aizuchi* based on the work of Goodwin (1981) and Kendon (1990). She found that Japanese speakers hold their recipients accountable for producing responses through the use of mutual gaze, much like Goodwin (1981) found for English speakers, but that unlike the English case, mutual gaze plus the production of *aizuchi* is possibly preferred over mutual gaze alone as a sign of listenership.

Aoki (2008), following Goodwin’s (1981) work on gaze and Kendon’s (1990) work on gesture analyzed the production of speaker head nods[33] in Japanese conversation and found that they are produced three different positions in relation to talk: turn-finally, at turn-internal prosodic unit boundaries, and during the production of a prosodic unit. She demonstrated how speakers use head nods to draw recipients’ notice to certain parts of their talk and elicit appropriate feedback at those points. Thus speaker head nods are a complex activity involved in the regulation and monitoring of recipient’s uptake and understanding. It follows that a*izuchi* responses are more strongly speaker-driven than previously assumed, and that speakers are

[32] Nakane (2007:21-22) criticizes Lebra (1987) and many other studies for lacking empirical evidence to back up their claims about silence in Japanese, and mentions many studies that use silence to present Americans and Japanese as complete opposites.
[33] Although speaker head nods are not usually considered backchannels, Maynard (1989,1997) discussed speaker head nods as a potential cue for backchannel responses.

actively creating recipients' opportunities to produce *aizuchi* via a variety of signals. Although *aizuchi* was not the focus of Aoki's (2008) study, her paper represents a positive step towards understanding *aizuchi* as part of a system of interactive actions occurring multimodally, rather than as a special class of behaviours that can be studied on their own.

Overall, then, the literature that focuses exclusively on Japanese backchannels or *aizuchi* is highly fragmented with little replication. The terms 'backchannel' and '*aizuchi*' carry a great deal of problematic baggage with them. In addition, what exactly they define remains unclear. It is my hope that this thesis will demonstrate that there may be alternatives to these problematic terms, but I believe that more work must be done before better terms can be put forward. For the remainder of this thesis I will simply use "recipient response" or describe specifically the behaviours in the data without categorizing them as backchannels or *aizuchi*. However, Aoki (2008), above, and some other papers which I will discuss below, are beginning to give new insights into backchannels and *aizuchi*. In this thesis I hope to expand on such research.

## 1.5 Other Approaches

Outside of backchannel/*aizuchi*-specific studies there are several other approaches that may have implications for the way we look at backchannels. Such studies address Schegloff's (1981) biggest criticism of backchannel studies in that they do not approach previously under-studied behaviours (i.e. backchannels and the like) that occur in conversation as if they are a separate (and separable) component of conversation, but rather as if they are an integral part that has simply been ignored. These papers have greatly influenced the way I approached the data in this thesis.

First, Charles Goodwin (1980, 1981), adopting a CA methodology, investigated the role of gaze in natural English conversation and found that mutual gaze is highly relevant to speaker restarts and pauses, and that gaze is one way that listeners display their listenership. In addition he demonstrated that speakers and recipients are on unequal footing when it comes to gaze use, with speaker gaze holding more weight during interaction, and recipients bound by it. He established two rules for gaze: first, that "speakers should attempt to obtain the gaze of their recipients during the course of a turn at talk (1980, 275)," and second, that "a recipient should be gazing at the speaker when the speaker is gazing at the hearer[34] (1980, 287)." Breaking these rules will result in speaker restarts, pauses, and other attempts to secure recipients' gaze. He also found that recipients may substitute for gaze other actions that display their listenership. Thus gaze is a part of a framework of a variety of interactive moves and possibilities.

Hayashi (2001), influenced by the work of Charles Goodwin and Adam Kendon (123) and working from a multimodal approach, looked at 'joint utterance construction[35]' in Japanese[36] and described in detail how participants use a combination of syntax, prosody, gesture, and gaze to negotiate the collaborative completion of another's utterance. Hayashi gives evidence for the syntactic, prosodic, and visual environments in which this activity takes place in Japanese, finding that rules for gaze and Gesture in Japanese appear to correspond with those described in English by Goodwin (1981), Kendon (1990), and others (122-123). I find that the environment in

[34] Goodwin (1980) uses the terms "hearer" and "recipient" interchangeably throughout the paper.

[35] In which one or more participants will actively take part in the completion of another's initial utterance.

[36]Most of Hayashi's participants were native speakers of the Kansai dialect, which is somewhat unusual as most studies focus on speakers of the so-called 'standard' (Tokyo) dialect. Hayashi does not note any issues due to dialect.

which joint utterance construction in Hayashi’s study resembles the PPU boundaries (page 12, above) described by Maynard (1989) in her description of backchannels, though much less specific. Thus there is evidence that prosodic phrase-unit boundaries with corresponding mutual gaze and gesture may be a highly active environment in which a variety of interactional behaviours occur, and that those identified previously as “backchannels” may be one part of a much more diverse system.

A presentation on preliminary findings by Ford, Thompson, and Drake (2009) on turn continuation in English conversation demonstrates how a variety of nonverbal actions, which they tentatively called “bodily-visual actions” or “BVAs” are “not only coordinated with talk but, interestingly, also extending beyond the verbal action” (1). Specifically, they demonstrate how a BVA performed by a speaker after verbal turn completion can extend his/her turn nonverbally. Two important issues brought up by Ford et al’s presentation are the difficulty in categorizing behaviours that are relatively non-specific; their meaning is dependent on the way they are combined with other actions and on the timing of production. They also mention the challenge of fitting multimodal data into existing notions of the turn-based framework when so much paralinguistic activity appears to initiate outside the rules of its linguistic counterpart. The authors do not refute the existence of the turn-based system, but suggest that it must be expanded or modified to account for observable evidence which has until now been passed over by investigators. The behaviours covered by Ford, Thompson, and Drake (2009) include facial expression, head nods, gaze change, and their various combinations.

In this vein, a recent study by Iwasaki (2009), using the CA framework, looked at how Japanese participants use verbal expressions, gaze, gestures, and facial expressions to create opportunities for co-production within a turn construction unit (TCU[37]). She calls these instances "interactive turn spaces (ITSs)." The environment of Iwasaki's ITSs are very specific places in which a speaker produces a noun or noun phrase which is separated from its following grammatical particle[38], and during that gap a recipient produces an intra-turn utterance. Iwasaki's analysis helps shed light on how Japanese participants interactively construct TCUs using a complex combination of verbal and bodily actions. The TCUs Iwasaki discusses do not fit into typical interpretations of what makes up a TCU in that there are elements of intra-turn interactivity. Like Ford, Thompson, and Drake (2008), Iwasaki (2009)'s findings suggest the need to re-evaluate or expand current notions of the turn based system to account for such intra-turn behaviour.

## 1.6 The Current Study

The above papers, which appear at first glance to have no direct connection to backchannels or *aizuchi*, actually deal with many of the behaviours that have been defined as such. It appears that a systematic global approach to the multimodal realities of talk may reveal more about the complex nature of backchannels/*aizuchi* than studies specifically looking at them. Table 1, below, sums up the general tendencies of previous backchannel/*aizuchi* studies, and how I intend to approach the topic in this thesis.

[37] The TCU is the fundamental unit of conversation analysis, based on speaker turns-at-talk. TCUs are bounded by turn-relevance-places (TRPs) in which speaker changes can occur.

[38] Japanese is a case-marking language with highly variable word order. Post-positional particles denote grammatical relations of constituents. Particles are often, but not always dropped in conversation. Utterances in Japanese conversation consist of phrase-unit-like chunks, often ending with a grammatical particle. Instances of a phrase and its associated grammatical particle being separated by a pause may be relatively uncommon.

Table 1: Comparison of this Thesis with Previous Studies

| Previous studies | This thesis |
|---|---|
| Tend to focus on arbitrarily selected features (ex. Vertical head nods) | Considers all behaviours as potential candidates before choose relevant features based on close observation of the data |
| Backchannels/*aizuchi* coded & categorized in relation to ongoing syntactic units (i.e. clauses and sentences) | Coding based on gaze and production within the multimodal environment |
| Focus on either verbal OR nonverbal; assumes non-coordination based on historical methodological preferences | Multimodal, inclusive, coordinated approach; assumes that verbal & nonverbal (and visual) work in conjunction |
| Categories pre-selected[39] | Categories based on observation of participant interaction |
| Narrow focus[40] | Broad focus |
| Heavily dependent on transcribed audio; limited use of video in actual analyses | Heavily dependent on source audio & video, transcripts for description only |

There are, of course, many exceptions to the claims I make in the 'Previous papers' column in Table 1, but I believe that as a general statement about the body of Japanese backchannel/*aizuchi* studies, it is accurate. The initial assumption that many papers on backchannels and *aizuchi* make, and that Schegloff (1981) criticises, is that backchannels exist as a set of similar, identifiable behaviours. Starting from this assumption, previous studies then select a behaviour or set of behaviours that they determine to be backchannels and then limit the analysis to those specific behaviours. Maynard (1989)[41] for example, chooses a specific kind of head nod (vertical) in a specific environment (near PPU boundaries). In most cases, these are the only behaviours that are coded along with talk, so the analyzed data is a small filtered set of what the

[39] Categories such as 'backchannel' or '*aizuchi*,' for example, are assumed to exist, despite the lack of consensus on what those words mean in the literature, or they are defined even before data is collected/analyzed, then further categorized functionally based on subsequent observation.

[40] Previous studies are narrow in terms of behaviour considered, environment, and the type of data used for analysis.

[41] I give Maynard (1989) as an example only because it has been so influential. The greater problem is that few studies have expanded beyond the type of behaviours Maynard initially chose to quantify.

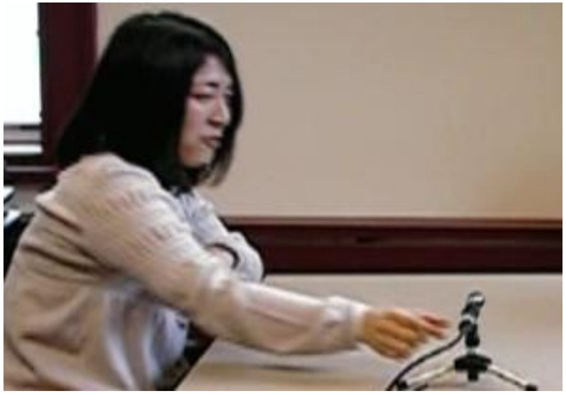
participants actually experience. Whether the category “vertical head nod” is relevant to the participant remains unknown. Do participants, for example, differentiate between clear vertical head nods and less clear vertical head nods, head shakes, or other head movements?

Of course there are good reasons to take such an approach. A quantitative study must define precisely what is to be coded; there is no room for ambiguity in numbers. Maynard’s (1989) selection of a specific type of head nods in a specific environment represents a reasonable start to the investigative process. But trying to categorize equivalent arm movements or posture shifts, which mechanically speaking have a much wider variety of possible motion, might be impossible to objectively code. For similar-looking/sounding responses that are directly related to the content of the talk-in-progress, such as facial expression and laughter, it would be very difficult to claim equivalence for two such tokens, and environmental similarities might not be present. These reasons probably contribute to the overall lack of replication and advancement in Japanese backchannel/*aizuchi* studies. However it is clear that this piecemeal approach has not furthered our knowledge of what backchannels/*aizuchi* really are (if they exist as a category at all) and how they work within the environment of talk.

My thesis consists of two parts: first I take a quantitative approach that builds on Maynard (1989), and then I take a multimodal approach to some specific examples of response behaviour. The quantitative study is actually two parts: the first part builds on Maynard (1989) by considering her observations of speaker-produced head nods as a possible factor in recipient response behaviour. The second quantitative section looks at speaker gaze as another possible factor, based on Goodwin’s (1980; 1981) observations from English. In the initial stages of the study, several attempts to organize the quantitative database were made, but once speaker gaze was coded, a clear pattern began to emerge, and the quantitative results confirm that speaker gaze is indeed a useful and relevant way to organize recipient response behaviour.  The results of the quantitative study show averages and trends but no detail; in addition the quantitative section cannot explain what exactly the participants are doing when they aren't following the trend. Therefore in the qualitative analysis section I take a look at some specific recurring examples of recipient actions to determine how the participants use gaze and bodily actions to achieve a variety of outcomes that are not accounted for in the quantitative analysis. I am interested in what the participants do and see, and what they notice and respond to, using as many of the resources I can that were available to the participants at the time of their interaction.

Traditionally, linguistics has been biased towards the written (Lidell 2005), and this bias has extended to the creation and use of transcripts as well (Ochs 1979). Rather than create confusing and error-prone multi-line transcriptions, I have chosen to reduce transcription to a minimum and use annotated video stills to demonstrate behaviour. While relying on video stills is still problematic, I think the adage "a picture is worth a thousand words" is very true: a still frame can capture much more that is going on, in a more easily understandable way, than any transcript can. I would also like to point out that language investigators tend to split behaviours into 'verbal' and 'non-verbal' categories, but I believe that 'visual' is a very important category as well. Visual elements need not be 'behaviours,' but simply 'what is seen or within sight,' and it is clear from my data that the speakers, at least, monitor their recipients as well as their environment for a variety of purposes. The picture below, for example, demonstrates how one speaker decided to use the microphone to represent a tree in order to create a visual node around which she could describe the spatial orientation of various people and objects in her narrative.

Figure 1: Utilizing visually available resources



This thesis is split into two parts, a quantitative section and a qualitative section. In the quantitative section I attempt to go a step beyond the behaviours that have been analyzed in quantitative studies thus far by expanding the analysis to include gaze and other nonverbal aspects of interaction. In the qualitative portion of this thesis I approach some specific instances in the data from a fully multimodal[42] perspective, with the assumption that every action is potentially relevant to participants' interaction. Inspired by Goodwin (1981), and supported by close observation of the data, the main focus of this study centres around units of mutual gaze in Japanese conversation and the behaviours by both speakers and recipients that occur both during and without gaze. Rather than attempt to use the terms 'backchannel' or '*aizuchi*' I take more or less a blank-slate approach. I assume that all participant behaviour is potentially relevant or connected to the ongoing talk, and closely observe participant interaction to filter out that which is non-relevant to them.

[42] It seems that the word 'multimodal' has been adopted to a significant degree in semiotics; that is not the theoretical/methodological approach I am taking here. Specifically I am looking at audio and video recordings and including in my analysis all manners of interaction (verbal and non-) as well as the environment as potential resources for the participants.

This study is based loosely on two previous and well known studies. The data collection process is a modified version of Chafe's *Pear Stories* process, while the focus of the study is based on Goodwin's 1981 examination of gaze and other extra-linguistic behaviours that occur during speech acts. The *Pear Stories* was one of the first major cross-linguistic attempts to examine spoken language using a well-defined experimental method of collecting audio[43] data. The resulting papers covered a wide range of topics on cognition as well as language difference (Chafe, 1980). Goodwin's CA-based (1981) study looks at more naturalistic conversation data and relies entirely on video data. Although his study is not linguistic by nature, an increasing number of linguists are using CA methodology as well as looking at data and behaviours that traditionally have not been considered within the sphere of linguistics. Likewise, the rise of Functionalism(DeLancey 2001) and the growing realization that language and behaviour cannot be explained so easily, if at all, in isolation (Steiner 1982:12), has led to an increasing number of studies that stray from traditional linguistics topics. Although this is very much a positive development, it is clear just from the issues within backchannel/*aizuchi* studies that methodologies need to be updated to cope with these topics.

In this thesis I make heavy use of video data to explore some paralinguistic behaviours in Japanese that have attracted the attention of linguists but remain somewhat of a mystery. In the following chapter I will discuss the data I collected, the methodology I used for coding the data, and some of the issues that I experienced during the process.

[43] Although it has been noted that some sessions were videotaped, the majority of papers based on the *Pear Stories* project appear to discuss only audio data.

# 2 Data

The main data for this study consists of five sessions of elicited narratives between Japanese native speaker dyads, based on a format modified from “The Pear Stories” (Chafe, 1980), in which participants were shown the Pear Film (1975) [44], a 6-minute-long film created by Wallace Chafe for the purpose of eliciting narratives for linguistic analysis. There is no clear plot and none of the characters speak, although there is sound. I give a brief summary of the Pear Film here to make it easier to understand the data in the next several sections.

## 2.1 The Pear Film

In the opening scene, a rooster crows, and an older man is seen on a ladder picking pears and placing them in a basket on the ground. There are three baskets in total. He climbs up the ladder once more and continues to pick pears. A boy on a bicycle appears, stops in front of the pear tree, seems about to take a pear, but then takes a whole basket, puts the basket on his bike and rides off. He is distracted by a young girl riding in the opposite direction and hits a rock, causing him to crash to the ground, spilling the pears onto the road. Three boys appear and assist him by gathering up the pears and helping him back up onto his bicycle. As the boy leaves, the three boys find his cap on the ground and give it back to him. As a show of thanks, the first boy gives each of the other three boys a pear from the basket. Meanwhile, the man has noticed that one of his baskets is missing. In the end, the three boys, munching their pears, end up walking past the very same pear tree and the old man from whom they were taken.

[44] The Pear Film may be viewed at <http://pearstories.org/>.

I chose the Pear Film because it was unlikely that the participants had seen it or even heard of it,[45] and also because of its uniqueness, in the sense that it has no clear plot, and the vast amount of literature based on it.

## 2.2 Participants

Japanese language is marked by a hierarchical system of politeness in which age and social status play a role. It also exhibits major differences in male/female speech norms, and a wide variety of local dialects. Politeness, gender, and dialect impact everything from vocabulary to syntax. For these reasons ethnographic information is provided here. It should be noted that the majority of Japanese language studies limit participants to those who speak the so-called "Standard" dialect of the Tokyo area, so this study departs from the norm. [46] All participants[47] were living in a medium-sized metropolitan city in Canada at the time of the study. All participants volunteered their time to participate. Participants ranged in age from 20 to 28. Table 2, below, displays the participants' ethnographic data.

[45] It turned out that one participant, Yuri, had heard the title before, but had never seen the film.

[46] One exception is Hayashi (2003), whose data is Kansai-dialect native speakers speaking (for the most part) the Standard dialect.

[47] I would like to thank the participants who volunteered their time and effort to assist my study and for allowing me the use of their likenesses in this paper. The names of all participants have been changed to protect their privacy.

Table 2: Participants

| Pair number | Name | Role | Dialect | Length of time in Canada | Relationship to partner |
|---|---|---|---|---|---|
| 1 | Kazu (M) | Speaker | Standard | < 1 year | Met once before |
|  | Hiromi (F) | Recipient | Kansai | 2 months |  |
| 2 | Aiko (F) | Speaker | Standard | < 1 year | First meeting |
|  | Yoshi (M) | Recipient | Standard | 2 months |  |
| 3 | Yuri (F) | Speaker | Standard | 5 years[48] | Acquainted |
|  | Azusa (F) | Recipient | Standard | 2 months |  |
| 4 | Akane (F) | Speaker | Kansai | 8 months | Good Friends |
|  | Nobu (F) | Recipient | Kansai | 8 months |  |
| 5 | Risa (F) | Speaker | Kansai | 8 months | Good Friends |
|  | Rintaro (M) | Recipient | Standard | 8 months |  |

As Table 2, above, shows, the speakers come from two main dialect groups, Standard and Kansai. For this study I did not specifically examine or control for the effects of dialect, gender, age, or length of stay in Canada.

## 2.3 Methodology

Participants were divided into pairs, each pair consisting of a "speaker" who was shown the Pear Film, and a "recipient" who had not[49]. Pairs were decided based solely

[48] Yuri had also lived in the United States for several years.

[49] The words "speaker" and "recipient" were not used when talking with the participants, although they did visibly adopt such roles once the exercise was underway. For simplicity, all subsequent uses of the words "speaker" and "recipient" in this paper will refer to these roles unless otherwise stated.

on participants' availability. All paired participants, except for Aiko and Yoshi, had met in some capacity before. Akane, Nobu, Risa, and Rintaro were good friends and arranged to participate together. The remaining participant pairs, Kazu and Hiromi, and Yuri and Azusa, had met at least once before but did not remember each others' names; they performed introductions during the data collection.

The participant's roles were decided in two ways; if one participant arrived before the other, they were shown the video and thus took the speaker role. If they arrived at the same time, I asked, "who would like to go first?" The participant who volunteered would then be shown the video and told that they would be asked to talk about what they saw with someone who had not seen it. They were shown the film alone in a closed room. Recipients were told that their partner was watching a film and that they were going to talk about it. The recipient was brought into the room immediately after the speaker had finished watching, and they were sat facing each other. They were both asked to continue talking until the recipient felt that he or she had understood the story, and were told that they were free to talk to each other as much as they wanted about whatever else they wanted to as well. At this point the participants could ask for clarification. One participant asked how she should approach the story telling, and was told, "it's up to you." All instructions were offered in Japanese, English, or both depending on the participants' preferences. The elicitation design is a departure from the original Pear Stories (Chafe 1975) setup, in which the recipient was an informed participant who was allowed only minimal participation with the storyteller, and the storyteller was more or less a one-way narrator, with no feedback or ability to converse with the listener.

Participants were recorded with a small digital video camera on a tripod placed perpendicularly to them roughly 2.5 meters away. The camera was not hidden but was placed outside the normal range of vision. The speakers were probably constantly aware of the camera throughout the session, because they would turn away from it when beginning a gesture, while in thought, or while performing a word search. A stereo audio recorder was placed on the table, between and slightly to one side of the participants in order to supplement the video camera's low-quality microphone. The resulting narratives ranged from just under 3.5 minutes to over 15 minutes in length (over twice as long as the Pear Film itself!). The interaction between speakers and recipients was very different as well, with some recipients actively conversing, and others remaining mostly passive.

## 2.4 On the use of elicited data

The genre of elicited narrative, specifically narrative talk based on the Pear Film, was chosen as the basis of this study for several reasons. In terms of discourse organization, the roles of speaker and recipient are well defined; although the participants were allowed to converse freely if they desired, they clearly assumed the roles of speaker/recipient during narrative sequences, allowing for analysis of verbal and nonverbal behaviour by role. This made the data perhaps easier (but not easy) to analyze. Naturally occurring conversation typically consists of fast, short turns at talk, and nonverbal behaviours such as gesture and gaze have been shown not to coincide exactly with spoken turns (Iwasaki 2009, Maynard 1989, Streeck 1993). Combined with the propensity for overlapping speech in Japanese speech overlap and the overall rapidity of the process it is sometimes difficult to clearly define whether head-nodding, for example, is being performed by a speaker or recipient. It is perhaps for these reasons that previous studies on Japanese backchannels tend to create criteria to only focus on very specific behaviours that can be clearly defined or demarcated from the surrounding action, rather than looking at those behaviours as one part of a complex interactive process. Narratives exhibit long turns at talk that allow for much easier analysis. It is definitely likely that natural conversation data will exhibit behaviours beyond what occurs in elicited narratives, but especially for preliminary analysis, coding accuracy was a great concern. Thirdly, because the narratives are all based on the Pear Film, a variety of possibilities for comparison and further investigation arise. Individual differences (and similarities) in, for example, vocabulary choice and gesture style can be analyzed based on points in the story. Additionally, elicited narratives (mostly audio-only) based on the Pear Film are numerous and have been collected from many languages, allowing for the eventual possibility of a large, multilingual corpus (Chafe 1980, Tannen 1993, Erbaugh 2001).

## 2.5 Coding and Analysis

Based on Goodwin's (1980, 1981) examples from English conversation and Hayashi's (2003) evidence that a similar system underlies Japanese conversation, I decided to sort the data based on points where the speaker made salient attempts to meet and break gaze with the recipient. As Goodwin (1980) observed, speakers make attempts to achieve mutual gaze with their recipients, and participants are expected to be gazing at the speaker at those points (287)[50]. I define gaze as one speaker looking directly at the other's face. Gaze can also be directed at another's body or hands or one's own hands, especially during gesture. 'Mutual gaze' involves the participants

[50] In this data set, recipients gazed at the speaker or the speaker's gestures more or less constantly. It is likely that mutual gaze in natural conversation is more complex due to the constant shifting between speaker/recipient roles between turns.

looking at each other. ‘Attempting to establish gaze’ means trying to attract another’s gaze toward one’s own face. ‘Meeting gaze’ means shifting one’s gaze to establish mutual gaze with another who is already gazing. ‘Breaking gaze’ is as it sounds; shifting the direction of gaze away from another, effectively ‘breaking’ a state of mutual gaze. Because gaze changes were observed to be made via quick, definite motions of the head, the majority of mutual gaze points and gaze breaks were relatively easy to code. However, participants can also use their eyes to make and break gaze, and this behaviour is more difficult to code, due to the fact that mutual gaze does not consist of constant staring, but rather regular scanning between face, hands, body, gestures, and even away for short periods. For this reason I only coded eye movements as gaze changes if they were lengthy (generally greater than 1 second) and were held in a constant direction (i.e. not scanning). If eye movement was used to break gaze, it was usually used to re-establish it, making this coding easier, but camera placement, video resolution, and even some of the participants’ hairstyles made it difficult to accurately determine gaze shift via eye movement in many cases.

How participants utilize gaze-related head and eye movements appears to be somewhat individual. For example, one speaker, Kazu, did not turn his head as far or as fast as other participants when making or breaking gaze, and would make glances at the speaker repeatedly with his eyes even when his head was turned away.[51]

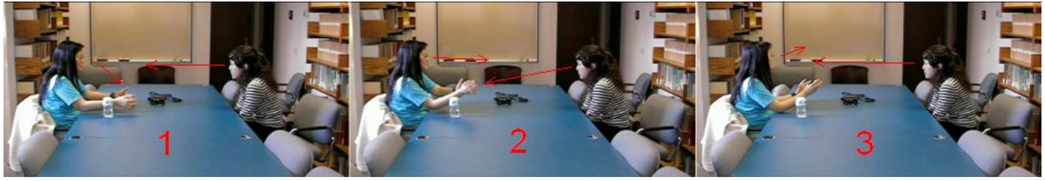
I also recorded clear instances where the speaker used his or her eyes to shift gaze quickly between his or her own gestures and the recipient. This draws the gaze of the recipient towards the gesture, following Streeck’s (1993) description of “objects of

[51] Kazu stated that he felt very nervous during his talk. His constant monitoring of the recipient and the smaller neck angles he achieved while gazing away compared to the other speakers could represent anxiety over the quality of his talk and a need to monitor the recipient for indications of trouble.

attention." Using gaze this way, gestures are made part of the ongoing talk, and refer to something the speaker said or is about to say.[52]

A gaze sequence that includes gesture is shown in Figure 2, below. In frame 1, Yuri (the speaker, left), has been in a state of mutual gaze with Azusa (the recipient, right), but momentarily looks down at her hands, which are starting to perform a gesture. In frame 2, Yuri has raised her hands off the table, making her gesture more salient, and looks at Azusa, who is now looking at Yuri's gesture. In frame 3, Yuri freezes her gesture and breaks gaze by looking away, while Azusa resumes gazing at Yuri's face. Yuri's gaze shift between frames 1 and 2, being very short and without a salient head movement (although there is a very slight change in head angle), were not counted as a gaze break/gaze meet. Her gaze shift in frame 3, however, accompanied by a salient head movement (and the freezing of her gesture), was counted as a gaze break.

Figure 2: Gaze unit sequence



Following these methods and criteria, all speaker gaze-initiation and gaze-break points in the selected portions of the videos were coded, along with actions that occurred nearly simultaneously or immediately after them, including gesture or changes in gesture, head nods, facial expression, salient changes in prosody, pause, laughter, posture and distinct phrase boundaries. Likewise, recipient responses, including gaze meeting, gesture, head movements, verbal responses, posture shifts that occurred near speaker gaze points were coded.

[52] Sometimes, the object of talk itself is explained via gesture only. Other times, the gesture may represent the quality of an object: its shape, size, height, location, etc.

Recording gaze initiation and gaze break was quite involved on its own. Accurately coding all actions that occurred around those points would probably be impossible. Through careful and repeated viewing of the video data, I attempted to record only those actions which seemed to be relevant to the participants. From my observations, relevant actions can be determined based on participant gaze direction and the relationship of sequential actions by speaker and recipient. There is also a behavioural trend that becomes quite obvious when observing speakers, which I will refer to as 'baseline behaviour.' It is clear that both speakers and recipients adopt relatively static stances in terms of posture, body and hand position, and facial expression for the majority of the talk. Large deviations from this 'baseline' appear to attract special attention and may be specifically used to attract it. This may be why gaze meets and breaks, for example, are usually performed with quick jerks of the head rather than smoother, more gradual neck movements or eye movement alone. Likewise, recipient actions that are relevant to the ongoing talk are similarly performed in quick, salient motions, while non-relevant actions, such as adjusting sitting position or posture to one that was more comfortable, were seen to be made very slowly, even unnaturally slowly, as if to avoid notice.

The method of behaviour identification, coding, and data sorting (i.e. the database) is the same for both the quantitative and qualitative analyses.

# 3 Quantitative Analysis

Because much work on Japanese backchannels is based on quantitative data, or at least quantitative claims, I decided to perform a quantitative analysis first, focusing on short responses and head nods. I then performed a second quantitative analysis sorting recipient responses by gaze. Although I was initially sceptical of the value of a quantitative approach, I found this exercise useful to compare findings with previous studies, especially since the speech genre of my data is different than many other studies, and also because, to my knowledge, nobody has done any quantitative work on the relationship between recipient responses and gaze in Japanese. In the end, the quantitative approach resulted in some very surprising and useful findings which guided the subsequent qualitative approach.

## 3.1 Recipient responses and speaker head nods

Acceptably quantifying all recipient behaviours turned out to be an impossible task for a couple of reasons. First, the variety of nonverbal actions made by recipients makes for an overwhelmingly large number of categories. Second, it is difficult to establish equivalency between two similar, but slightly different movements for the purpose of categorization, since they are often in response to something specific in the speaker's talk. For this reason, I decided to follow Maynard's (1989) categories and code only clear examples of vertical head nods for this part of the quantitative analysis. Considering the differences between the data used in Maynard (1989) and this thesis (naturally occurring conversation vs. elicited narrative, respectively), there is value to this decision. Based on Maynard's (1989) observation that speakers also make head nods, and Maynard's (1990) quantified results of speaker head nods and recipient responses, I decided to include speaker head nods as a category as well. In the second part of the quantitative section I will add gaze to the equation, which has not been done before.

Vertical head nods consist of quick up or down movement away from the normal rest position of the head, which is more or less horizontal. Head nods were observed to be produced in the following forms: down[53], up-down, down-up-down, and down-up. Head nods that ended in an upward movement (i.e. down-up) or head movements that did not have a down motion at all (up only) were not counted because they were determined to be functionally very different from the other types[54]. Head nods that were produced in succession, with no pause between them, were counted as a single nod. All kinds of verbal responses were counted, including laughter and answers to direct questions and questions relating to the speaker's talk[55]. Verbal responses in this data were mostly limited to short utterances, with *un* being by far the most common.

My initial plan was to select 4 minutes of data from each pair starting from the time the speaker started describing the Pear Film. However not all the conversations turned out to last for 4 minutes, and I decided to stop coding at less than 4 minutes for one conversation because the participants held a lengthy period of conversation-style talk in the middle of the exercise. Since that deviated from the rest of the selected data, I decided to simply stop at that point[56]. To make the database, I first used the

[53] Of course, participants must raise their head to return to a horizontal position, but the return motion is often slower than the down motion, and the head does not go past the baseline on its way up.

[54] Upward head motion will be the main topic of the qualitative section.

[55] Questions are sometimes not considered to be in the same category as 'backchannels' and thus some studies do not count them, for example Sugito (1989) and Tanaka (1999). However because there were less than 10 instances I decided to include them.

[56] I could have reduced all the selections to 3 minutes each, but I was unwilling to simply throw out the coding I had performed. Furthermore, the quantitative data is not the main focus of my paper and I am not making any claims based on averages here, since my methodology is

audio/video file from the camera. After watching the video data, I listened to the digital audio data that was recorded separately with a dedicated audio recorder and filled in the database with utterances that were not audible in the video file. I estimate that about 30% of recipients' responses were not audible in the video files, but were very clearly audible in the separately recorded audio file[57]. Needless to say, a 30% difference would have a drastic effect on the accuracy of the quantitative analysis. The table below condenses the quantitative observations.

**Table 3: Quantitative Results of Recipient Responses by Type**

| 1. Speaker name | 2. Recipient name | 3. Selection length (m:ss) | 4. Responses per minute | 5. Mixed nods + verbal responses[58] | 6. Verbal only responses | 7. Nod only responses | 8. Total responses | 9.Participant responses after speaker nods |
|---|---|---|---|---|---|---|---|---|
| Kazu | Hiromi | 4:17 | 17 | 33 | 4 | 34 | 71 | 19 |
| Aiko | Yoshi | 3:14 | 16 | 21 | 8 | 23 | 52 | 30 |
| Yuri | Azusa | 4:19 | 33 | 45 | 2 | 94 | 141 | 83 |
| Akane | Nobu | 3:23 | 22 | 57 | 8 | 11 | 76 | 70 |
| Risa | Rintaro | 3:38 | 14 | 22 | 15 | 13 | 50 | 47 |
|  | **Totals:** | **18:51** | **20 (average)** | **178** | **37** | **175** | **390** | **249** |
|  |  |  |  |  |  |  |  | **Total Speaker nods: 270** |

The pseudonyms of the participants and their roles are displayed in columns 1 and 2 on the left: shaded pink for speakers and green for recipients. For ease of understanding, all pink shaded sections are associated with the speakers, and all green sections are

admittedly not as strict as, for example, Maynard (1990;1997), nor do I consider my data set large enough to support any general claims.

[57] Because the most common response, *un*/*n*/*m*, can be produced without moving the lips, there is absolutely no clue that such responses are being made without high quality audio recordings to supplement the video camera's audio track. The same issue may affect other studies that rely solely on consumer video equipment or camera-based microphones as well.

[58] Several kinds of verbal responses occurred, including laughter, although the vast majority consisted of *un* and its variants, *n* or *m*.

associated with the recipients. Column 3 shows the duration of each interaction I analyzed; the total is 18 minutes 51 seconds. Column 4 displays the average responses (both verbal and head nods) per minute (calculated by dividing the total responses in column 8 by the selection time in column 3) to give a sense of the overall response frequency[59], with the overall average for all five recipients (20/min) at the bottom. A single response could consist of:

1. A simultaneously produced combination of a verbal utterance + head nod (column 5);
2. A verbal utterance only (column 6); or
3. A head nod only (column 7).

Verbal utterances and head nods that were produced more or less simultaneously were counted as a single 'mixed' response in column 5. When a recipient produces a mixed utterance the initiation of the nonverbal head nod typically precedes the corresponding vocalization, with the vocalization typically beginning as the recipient's chin reaches its lowest point. It is very unlikely that the verbal/nonverbal components represent separate responses, but rather are produced as a single response occurring across two modalities. An example will be given after Table 3 is fully described.

Column 8 shows the total number of responses by each recipient (the sum of columns 5, 6, and 7). Column 9 counts the total number of recipient responses (making no distinctions between verbal, nonverbal, and mixed) that were produced *only after the speaker nodded his or her head*, with the total number of speaker head nods noted at the bottom, in pink. I chose to add this count to the study based on Aoki's (2008) finding that speakers' head nods have an effect on recipients' production of responses.

[59] Because speakers' speaking speed and the overall organization of their talk may vary, 'responses per minute' is not an objective measurement of recipients' response behaviour.

As can immediately be seen, recipients responded immediately after a speaker head nod 249 out of 270 times; this will be discussed in more detail below.

A surprising finding was the apparent preference recipients had for responding via head nods or head nods with an accompanying verbal component (columns 5 and 7), rather than responding only verbally (column 6). Participants produced a head nod in 353 out of 390 total responses, or 90.1%[60] of the time. This is higher than Maynard's (1989) figure of 63.15%. Why there is such a discrepancy is unclear, but what is clear is that overall, head nods appear to be an important constituent of response behaviour among the participants in this set.

The effect of speaker head nods on the production of recipient responses seems to be great: speaker head nods preceded recipient responses 249 times out of 390, or 63.8% of the time (column 9). This is much higher than the 38.08% recorded by Maynard (1990). I followed up on this by counting how many times recipients did *not* respond after a speaker made a head nod; the total was only 21 times[61] out of 270, or 7.8% (bottom of column 9). Looking further, I found that the speakers were not always gazing at the recipient while nodding; as I will demonstrate in the following section, a lack of a state of mutual gaze could reduce the impetus for recipients to make a response[62]. I did not differentiate between recipients' verbal and head nod responses in the context of speaker head nods, so I cannot compare figures directly with those in

[60] Azusa made substantially more head nod-only responses than any of the others, but removing her data from the calculations still results in a very high 85.9% of all responses having a head nod component.

[61] I only quantified the total 'unanswered' speaker head nods, rather than quantifying the number for each speaker.

[62] Because my database is based on durations when the speaker is gazing at the recipient, speakers head nods made while looking away were mostly not recorded.

Maynard (1989)[63], but these initial results are more fruitful than expected; I would like to follow up on this with a larger sample.

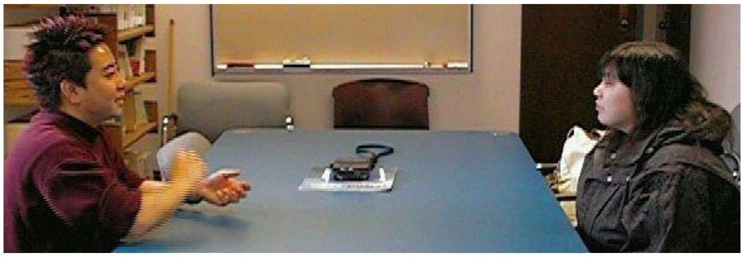
The table demonstrates very clearly that recipient response is highly interactional and often speaker-driven (through the speaker's production of a head nod). As well, the table shows that recipient responses vary wildly depending on the individual. Azusa, for example, responded with a head nod-only twice as often as she did with a head nod-verbal combination (94 versus 45 times), but made a verbal-only response only 2 times out of the 141 total responses she made. Rintaro, on the other hand, produced verbal-only and head nod-only responses nearly equally (15 versus 13 times), and though he too showed a preference for head nod-verbal combination responses (22 times), the differences are minimal compared to Azusa. These numbers suggest a corpus for the purpose of unbiased quantitative analysis of each type of response would have to be very large to overcome individual difference alone; much larger than any study has attempted thus far. I don't think that a larger corpus would necessarily bring better results in terms of the overall trend for recipients to produce some kind of response after a speaker head nod, however. And individual difference does not have to be a limiting factor if followed up with a quantitative analysis, as I attempt in this thesis. The effect of other factors, such as age and dialect remain unknown. As for gender, Tanaka (2004) cites four papers that discuss the relationship between *aizuchi* and gender. The consensus among them is that females produce quantitatively more *aizuchi* than men overall (178), and that individual *aizuchi* behaviour may vary depending on the gender of the interlocutor. In her own data, however, Tanaka finds "no striking difference"

[63] Maynard (1989) states that 38.05% of recipient backchannels occurred in the context of speaker head movement 174). Because Maynard differentiates backchannels from head nods in other parts of the paper, I have to assume that this figure does not include head nods.

between male and female production of *aizuchi* (ibid). Clearly more investigation is necessary into the effects of gender. Since the results show that speaker head nods occur before most recipient responses, there is also the possibility that the way the speakers produce their narratives affects the number of opportunities recipients have to make responses as well as the type of response the recipient makes. For example, if the speaker talks without producing many head nods, the recipient may not feel it necessary to respond. If the speaker produces long stretches of unbroken talk with few instances of pause, a recipient may produce more head and fewer verbal utterances in order to avoid overlapping talk. Likewise, speakers' skill at storytelling could greatly affect the recipient's interest and, by extension, the number or type of responses the recipient produces. The relationship between these kinds of speaker actions and recipient responses is unknown and requires more investigation; how to measure these factors is also an issue.

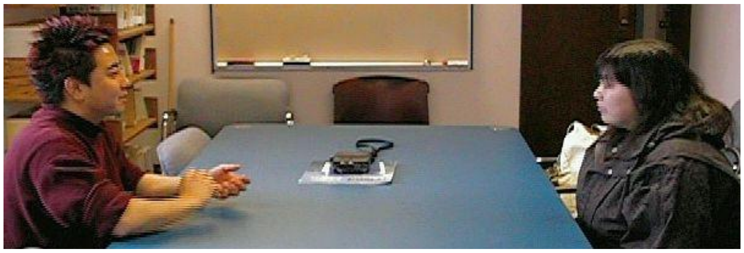
Before moving on I believe it is necessary to demonstrate how I coded head nods, which can be produced in a variety of ways, and especially the combined verbal utterance + head nod category, which I propose represents a single response produced across two modalities. Below is a typical example. The speaker, Kazu, is on the left, and the recipient, Hiromi, is on the right. Kazu is coming to the end of his clause, *nashi o totteiru aida ni* "while (he) was picking the pears" (Line 1) As Kazu starts to say *ni*, Hiromi closes her eyes[64] and begins to initiate a head nod (coded with capital "N" in Line 2) as shown in the still frame below.

[64] Closing one's eyes is not requisite for producing a head nod in Japanese; it is just incidental to this example.



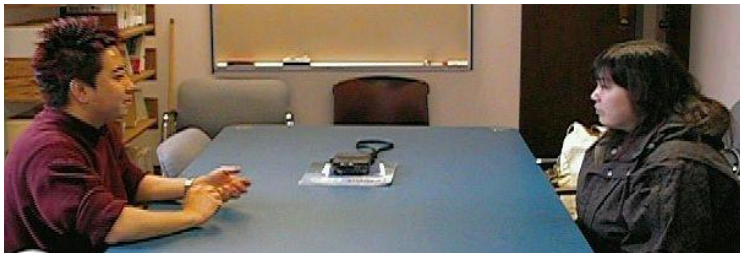
```
1. Kazu:   nashi o   totte  iru   aida ni, (1.0)
           Pear OBJ pick   -ing  while
           "while (he) was picking the pears"
                                       [
2. Hiromi:                             N (initiated)
```

The following frame shows Hiromi's head position at the lowest point of her head nod. Only about 1/29[65] of a second, i.e. a single frame, has elapsed since the previous image.



```
3. Hiromi: N (lowest point)
```

Finally, one frame later, Hiromi, with her head still down, opens her eyes and says *n* 'uh-huh.' Kazu's voice has stopped by this point.



```
4. Hiromi: n
          "uh-huh"
```

Thus 'mixed' responses exhibit a slight difference between the initiation of the head nod and the initiation of the verbal response. In the case shown above the difference is minimal at only 2/29 of a second, but the delay is quite variable, up to around half a

[65] Each frame of NTSC video (the standard video format for North America) represents roughly 1/29 s.

second, and possibly more. If the recipient kept his or her head low while responding verbally, despite an obvious gap between the initiation of the head nod and the verbal response I counted it as a single response. Other instances in which there was a significant gap between the production of the nonverbal and verbal response were dealt with on a case by case basis, taking into account the content of the speaker's ongoing talk when each recipient response occurred. There were only one or two instances in which this occurred, so the effect of subjectivity should have a negligible effect on the final counts. In addition to producing responses simultaneously, recipients can produce responses iteratively, for example by producing three head nods in sequence or by repeating *un* "uh-huh" several times in a row. These were counted as a single instance (as in Maynard 1989) unless an identifiable pause occurred between them. Likewise a stretch of laughter or chuckling was counted as a single response unless there was a noticeable pause between instances. Cases in which a series of head nods were followed immediately by *un* were also counted as a single mixed case. Counting head nods and verbal responses completely separately would result in obscuring the mixed cases in which they were produced together and would create an unbalanced comparison in which responses produced nearly simultaneously but across the verbal and nonverbal modalities would be counted twice. I believe there is sufficient evidence that a recipient response which includes both a verbal component and a head nod should be considered to be a single response that just happens to occur in two modalities[66].

[66] The division of verbal and nonverbal is another example of linguistic bias stemming from textual analysis, according to Liddell (2005:118).

## 3.2 Results for the Quantitative Analysis of Gaze

Gaze was coded based on the criteria specified in section 4.5. I created a database which uses the duration of speaker-established mutual gaze as its main unit, based mainly on Goodwin's (1981) description of gaze in English, but also through trial-and-error-based attempts to find a way to code the data in a way that reflected the participants' interactions. There were also clues in Hayashi (2003) that gaze is an important consideration in other complex multimodal interactions such as joint utterance construction; I had observed evidence in my corpus that recipient responses were probably being negotiated in a similar way. Each unit begins with mutual gaze (or a speaker's attempt at establishing mutual gaze) and ends with the speaker breaking gaze[67]. The duration of a gaze unit is entirely up to the speaker; they ranged from less than 1 second to over 20 seconds in length. I coded the points in the video files when each speaker achieved or attempted to achieve mutual gaze with their recipients, as well the times and descriptions of all recipient responses[68] that occurred during the period of mutual gaze. Other speaker actions, if any, that occurred immediately before the recipients' responses occurred were also coded,[69] to account for the possibility that speaker syntax, pitch, head nods, or any other salient moves could be prompting recipient responses[70]. All recipient responses that occurred outside of gaze units were

[67] Recipients can break gaze as well; these points were also coded. The qualitative section will examine how recipients who break gaze affect the speakers' ongoing talk.

[68] I did not limit the count to 'vertical' head nods in this section, nor did I distinguish between verbal nonverbal responses. For this reason, the numbers are slightly different than those in Table 3, and they cannot be directly compared. However it is easy to extrapolate the number of extra responses in Table 4 by subtracting the values from Table 3.

[69] I noted down words, syntactical information, prosody that stood out (rising, lengthening, cut-offs, etc), gestures, head nods or other salient bodily motions, and facial expression.

[70] This data turned out not to be necessary for the quantitative analysis of speaker gaze, but was very helpful in the qualitative analysis. With some modification to the database, this type of data could be very useful for a future project looking at the relation between syntax and gaze.

also recorded, although they turned out to be very rare, as will be discussed below. The data turned out to be very useful for the qualitative analysis, however. Finally, the point where a speaker broke mutual gaze from the recipient was recorded. The duration between the speaker establishing mutual gaze and breaking gaze is what I call a 'gaze unit.'

The process of recording gaze units was very time consuming and difficult. Luckily, however, digital video format makes it easy to jump back, slow down, automatically repeat, and even move frame by frame. Unfortunately most commercial playback software does not provide time readouts in increments smaller than a second, and I noticed that time readouts can be slightly differently between different software packages i.e. the point in time from the start of the video displayed as "03:04:25" in (mm:ss:ms) in one package could appear as "3:04:14" in another. I therefore sacrificed accuracy at the millisecond level for consistency at the second level. If more accurate timing was required (for the qualitative section, especially), I used a different software package to measure the duration of the action, without regard to the point in time as registered by the software's clock.

Sorting the data based on gaze points and durations revealed very interesting results. The table below covers some general findings from the data.

Table 4: Quantitative Results for Speaker Gaze

| 1<br>Speaker Name | 2<br>Recipient Name | 3<br>Length of selection (m:ss) | 4<br>Number of Instances of Mutual Gaze | 5<br>Total gaze duration[71] | 6<br>Recipient responses[72] during speaker gaze | 7<br>Recipient responses without speaker gaze |
|---|---|---|---|---|---|---|
| Kazu | Hiromi | 4:17 | 39 | 3:04 | 76 | 1 |
| Aiko | Yoshi | 3:14 | 34 | 2:01 | 64 | 0 |
| Yuri | Azusa | 4:19 | 49 | 2:13 | 140 | 10 |
| Akane | Nobu | 3:23 | 38 | 2:00 | 83 | 5 |
| Risa | Rintaro | 3:38 | 30[73] | 2:11 | 56 | 5 |
| **Totals:** | | **18:51** | **190** | **11:29** | **419** | **21** |

Columns 1, 2, and 3 show the speakers' pseudonyms, recipients' pseudonyms, and the duration of the data selected for coding, respectively. This selection is the same as was used for quantifying recipient responses in section 5.1. Column 4 refers to the total units of mutual gaze coded in the database, each starting with the establishment of mutual gaze and ending with the speaker breaking gaze by looking away. Kazu, for example, established and broke gaze 39 times. In total, I looked at 190 gaze units (bottom of column 3). Column 5 displays the total time each speaker spent gazing at the recipient. Thus, Aiko spent 2 minutes 1 second gazing at Yoshi in total of 34 gaze units

[71] The duration was rounded to the nearest second for each instance of gaze and are provided to give a general idea of how much time the speaker spends looking at the recipient.

[72] Recipient responses include both verbal and nonverbal responses. Verbal responses were overwhelmingly "un" or "n," although Azusa produced only "hai."

[73] Counting Risa's gaze units was difficult due to her hairstyle and her tendency to face slightly away from the camera while talking, obscuring her face. This may partially account for the low value.

during the 3 minute 14 second selection. The last two columns are the most important in terms of the analysis. Column 7 gives the number of recipient responses made while the speaker was gazing at the recipient, and column 8 gives the number of recipient responses made while the speaker was gazing away from the recipient.

The first finding is that recipients tend not to give responses unless the speaker is gazing at them. Out of 419 responses, only 21 (5%)[74] of them occurred when the speaker was not gazing at the recipient. A single speaker, Azusa (whose response behaviour also differed from the others in the previous section), was responsible for producing just under half of these 'anomalous' responses, but there is a good explanation. Azusa is acquainted with Yuri and knows that she is older and a Japanese language instructor, making Yuri her senior. Evidence that she is adjusting her behaviour to this perceived social hierarchy comes from Azusa's exclusive use of the more polite "*hai*" (compared to the more familiar "*un/n/m*" used by the other recipients). This also explains her greater overall attention to the speaker in terms of the number of responses produced: 150 total responses, double or greater than all other participants, despite Yuri gazing at her for the shortest relative duration[75]. For quantitative studies on backchannels, then, perceived social level may have a prominent effect on both the number and type of backchannel response. It is also possible, however, to explain Azusa's behaviour simply as individual difference. Another possible factor is that Yuri's production of Japanese is influenced her L2, English; Yuri has been in Canada the longest

[74] This number is, suspiciously, the same as the 21 instances where recipients did not respond to speaker head nods in Table 3 in the previous section. Because of the slight differences in methodology between this section and the previous section, I will have carry out another analysis to see if they correspond.

[75] It is also possible that Azusa's high level of demonstrated attentiveness via her responses reduces Yuri's need to monitor her via lengthy periods of gaze.

of all the participants and speaks English at near-native competence, and that this may have an effect on how Azusa produced her responses.

In any case, the data shows that recipient responses are highly dependent on the state of the speaker's gaze. Because gaze units are of varying length and a lot of things happen within their boundaries, this is not really a 'smoking gun' in terms of recipient response production. It does, however, suggest that gaze units may be a defining environment for most interactional events. A quantitative analysis on a larger corpus of conversational data is an obvious next step.

Table 4 shows that the speakers spend quite a bit of their talking time gazing at the recipient, anywhere from 50% to 75% of the time, depending on the speaker. The average time speakers spent gazing at recipients is 61%. Gaze duration for each instance is highly variable, however, ranging from less than one second to upwards of 25 seconds at a stretch. Gaze duration is, of course, entirely up to the speaker. I did not carry out a systematic investigation on the relationship between gaze and the content of speaker's talk, but casual observation suggests that speaker's gaze duration is longer when making simple descriptions, and shorter when describing more complex scenarios, such as spatial relationships. For example, when speakers were describing the characters' appearance or the general setting of the Pear Film, they looked at the speaker for long periods and glanced away for short periods. When they described spatial relations or characters' movements, the speakers spent more time looking at their own gestures than at the recipient, who they monitored only briefly. This may reflect the greater concentration required to present the narrative (which is not entirely coherent in the Pear Film) in a coherent way for their recipients, at the expense of monitoring them. Further investigation into this observation is necessary.

Gaze was also seemed to be more regular (and gestures few) during the initial minute of the narratives. The initial period may be a “getting to know each other” phase in which the speaker closely monitors the recipient’s behaviours in order to learn their individual traits (since the evidence above demonstrates that recipient response behaviour may be highly individual), or could reflect the simplicity of narrative at the beginning. Again, further investigation is necessary.

Recipient gaze times were not calculated because recipients spend the majority of their time gazing at the speaker or the speaker’s gestures, and are expected to do so (Goodwin 1980). This is not to say recipients’ gaze is unimportant. Instances where recipients broke gaze with the speaker or were not looking at the speaker when expected were quickly treated as a trouble source and resulted in some kind of repair or negotiation at the next available opportunity, determined by the speaker. Thus a change in recipient gaze from the expected baseline is probably important, while duration is not. Examples of how recipient gaze and other actions affect speaker talk will form the basis for the qualitative analysis section.

## 3.3 Summary of Quantitative Results

The main findings of the quantitative sections were as follows. In the first quantitative section, based on recipient responses, recipients were found to prefer responding with head nods or a combination of a head nod with a verbal response, rather than a verbal response alone. Second, recipient responses tended to follow speaker head nods 63.8%[76] of the time, which corroborates Aoki’s (to appear) claim that speaker head nods “invite immediate responses from recipients” (1). In the second

[76] As stated earlier, only speaker head nods made while they were gazing at the participant were quantified. It is unknown whether speaker head nods made while the speakers were looking away contributed to recipient production of responses in the 21 times they occurred.

section, gaze was considered, and speaker-initiated mutual gaze was found to be highly relevant to recipient production of responses; recipients produced very few responses in the absence of mutual gaze. Speakers can make numerous head nods during a gaze unit as can be inferred by comparing the numbers: 270 head nods in 190 gaze units. The very small number of responses made when the speaker was not looking at the recipient (21 out of 440 instances) suggests that the state of speaker gaze has a greater impact on overall recipient response production than speaker head nods, although once gaze is established a speaker head nod may act as a strong indicator that a recipient should respond at that point. In other words, a recipient produces very few responses if the speaker is not gazing at him/her, but once gaze has been established, recipients tend to perform most of their responses after speaker head nods.

My hope is that the quantitative section, despite its limitations, has demonstrated clearly that recipient responses may be heavily dependent on at least two non-linguistic factors that have been under-considered in previous studies: speaker head nods and gaze. It is likely that more factors will be discovered. In the next section I will perform a qualitative analysis to demonstrate more clearly the importance of gaze and nonverbal actions in speaker-recipient interaction.

# 4 Qualitative Analysis

In the quantitative section, speaker gaze and head nods were found to be factors in recipients' production of responses. Gaze has previously been shown to have a regulatory effect on conversation in English (Goodwin 1980; 1981) and Japanese (Szatrowski 2002). Goodwin (1980; 1981) gives some examples of what happens when recipients fail to return gaze at appropriate times in English, namely pauses and restarts, but little has been done to investigate how recipients utilize speaker gaze as a tool for negotiation for their own purposes. The quantitative results in the previous section of the thesis showed that recipients do not respond in an expected way 100% of the time, and the database showed that they are sometimes active outside of speaker gaze units. In other words, recipients were not just passive listeners but active participants in the ongoing talk, and I found a pattern of certain types of recipient responses that appeared to have a great impact on the speakers' talk. In short, I discovered that both speakers and recipients use gaze in a variety of ways. Speakers use gaze to monitor their recipients and to create spaces in which recipients can (or are expected to) respond. Recipients, who are expected to gaze mainly at the speaker (Goodwin 1980), were found to break or avoid meeting gaze with the speaker for specific purposes, resulting mostly in speaker pause and repair (similar to Goodwin 1980), but also sometimes resulting in the recipient gaining a turn at talk. Although the outcomes are very different, they are all tied to the participants' use of gaze as a resource.

Using annotated video stills and a fully multimodal approach in which all speaker and recipient behaviours were first assumed to be potentially relevant to the talk at hand, I will present some specific interactive behaviours that have not typically been considered to be backchannels or *aizuchi*[77], even though they occur in the same environment. Specifically I will be looking at instances where speakers moved their heads or looked up (in contrast to the downward-trajectory head nods in the quantitative section), resulting in several different outcomes. I hope to show that backchannel behaviours, as they have been introduced by previous studies, should not

[77] Many of the features I look at in this section were actually considered as potential candidates in Yngve (1970), although that paper did not go further than listing them. Many of Yngve's suggestions, especially those involving nonverbal behaviour, seem to have been overlooked or ignored in later studies. Schegloff (1982) also mentions many potential nonverbal candidates but they do not enter his analysis.

be considered independent units, but rather a small part of a synchronous/sequential multimodal system that is constantly being monitored and adjusted by both speakers and recipients. In the quantitative section above, I chose portions of the data for analysis. In this section, however, I pick some of the more obvious examples from the entire corpus. The sequences I am describing are not quantifiable since they depend on the participants' actions and reactions to the content of talk, all of which is constantly changing and dependent on a wide variety of factors. I hope to demonstrate that there is a system at work, and that what have been previously called backchannels are actually a part of this interactive system of behaviours. I also argue that they should not be separated into constituent parts for the purpose of analysis or categorization based on function, since each part's function is bound to other actions that happen with them, both simultaneously and in sequence. Table 5 summarizes the main points of the following qualitative section.

Table 5: Recipient Manipulation of Speaker Gaze

| Recipient Behaviour | Outcome |
|---|---|
| 1. Recipient gazes back and/or provides feedback when gazed upon by speaker | a. Speaker continues turn at talk |
| 2. Recipient purposefully breaks or avoids speaker's gaze with upward head movement | a. Speaker pauses until gaze is returned<br>b. Gaze is not returned; recipient gets/takes a turn at talk |
| 3. Recipient makes salient posture change, avoids giving feedback, and looks up | a. Speaker performs repair (specifically an increment) |

The following sections will look at how speakers and recipients use gaze as a tool to negotiate a variety of outcomes, including turn change. After demonstrating a 'normal' instance in which the speaker's establishment of mutual gaze leads to a recipient response, which is the most common behaviour at gaze points in the corpus, I will look at several cases (described briefly in Table 5, above) in which recipients use gaze along with other bodily actions that are not typically considered in backchannel/*aizuchi* studies. I will start with an analysis of transcribed data, and then perform an analysis of the video. I do this for two reasons: first, to present the data in a way typical of previous, quantitative analyses of backchannels; and second, to demonstrate how much more is revealed through the use of video data.

The transcripts are created using numbered lines for reference. Because the video stills reveal a great deal of nonverbal behaviours that are produced simultaneously with speech, the line numbers in the video stills may not correspond with those in the transcripts that focus on spoken data. Line breaks may occur due the margin limitations or because significant recipient action occurred during the speaker's talk; for the purpose of clarity I have broken up the turns at talk into workable amounts. I use tabs to line up recipient responses with speaker talk in the places they occurred. Head nods are noted with a capital "N" and they are placed on a separate line if talk also occurs. Descriptions of other relevant actions are enclosed in brackets as follows "<action>". Continuing actions are denoted with a string of dashes and their completion marked with an X: <action-----------------------------------X>.

Before looking at the data itself, I would like to introduce some of my initial observations on the use of gaze as an interactional component of Japanese conversations.

## 4.1 Initial Observations on Gaze and Interaction

The most active parts of gaze seem to be gaze initiation and break points, at which a variety of speaker/recipient interactional actions occur. Speaker gaze initiation points slightly precede the production of phrase endings, which Maynard (1989) noted, are the environment of the production of backchannels[78]; thus they are an active environment for speakers to monitor recipient reactions, listenership, and understanding, and, as Szatrowski (2007) notes, for prompting *aizuchi*[79] and for recipients to produce such behaviours. Gaze points could thus be considered to be linked to utterances at a syntactic level[80]. As determined in the quantitative section, above, most recipient responses occur after speakers establish gaze. It is therefore likely

[78] Sacks (1992) similarly notes that in English, 'uh-huh' "ties to some last utterance, clause, phrase" (746).

[79] Szatrowski (2007:122) discusses "utterance endings," but describes speaker behaviours such as "direct gaze," "head nods" and "sound stretches," all of which can occur at phrase boundaries in Japanese.

[80] McNeill (1985) discusses the possibility that gestures and speech share a computational component. Gaze, however, is not gesture, and it is not clear what, if any, speech-related computational component exists within gaze, or if the timing of speaker gaze is based on the self-monitoring of the internal computational processes of speech, or something else entirely.

that recipients use speaker gaze to help time their responses, although gaze cannot be the only factor, since non-face-to-face modes of talk are completely free of gaze. Most likely, recipients monitor a variety of clues during speaker talk, one of which is gaze. The frequency of recipient responses that occur near gaze-unit boundaries suggest that speaker gaze could be considered one factor of projection[81] in Japanese.

Points where speakers break gaze do not seem to correspond with phrase boundaries but do often occur with gesture initiation, as in Streeck (1993). In terms of speaker behaviour, then, breaking gaze just signals the continuation of talk. But recipients sometimes take a turn at talk at the exact moment a speaker breaks gaze, as will be discussed below. Speaker gaze, then, is a visual tool used by both speakers and recipients in very different ways. Recipients are expected to meet speaker's gaze (Goodwin 1981), but they don't have to, which can lead to several outcomes. These outcomes will be the basis for the following sections.

Participants' behaviour, especially that of the recipients', was quite static in terms of overall posture and positioning throughout the interaction. The way they positioned their bodies at the beginning was held through to the end. If one was to watch each video at very high speed, where individual frames lose their effect, there would appear to be little movement. I call this tendency 'baseline behaviour' and found that it was useful for analysis. Large deviations from participants' baseline behaviour, for example gesture or posture shift were points of heightened interaction. Speakers noticed and attended to recipients' movement away from the baseline. I will discuss baseline behaviour in greater detail in a separate section below.

[81] For discussion of projection in English based on analysis of audio, see Schegloff (1987).

## 4.2 Speaker Gaze and Recipient Responses

This first example will demonstrate how speaker gaze creates an interactive space for recipients to respond. This example could be considered a typical example of a recipient response in Japanese. In the transcript below, Yuri (YU) is telling Azusa (AZ) about how the young boy has put the basket of pears back on his bike after falling down and is about to ride away from the three boys who helped him up. First I will show the audio transcript, and then follow up with the corresponding video stills.

In line 1, Yuri explains that the boy took the basket and left, to which Azusa responds with *hai* in line 2. Azusa's response overlaps with the continuation of Yuri's talk in line 3.

### 4.2.1 YU-AZ Selection

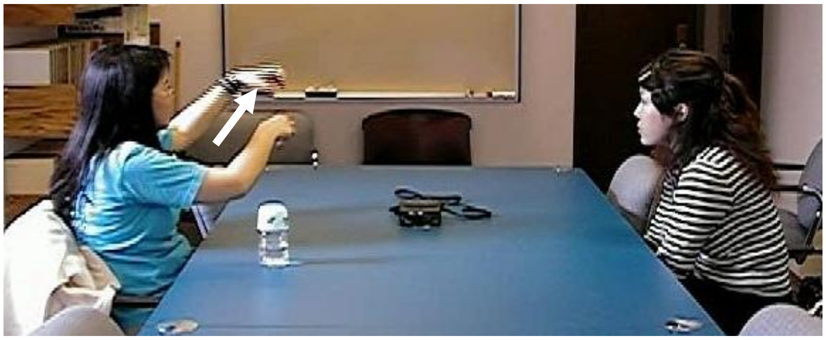
```
Data: [YU-AZ] 3:50-3:53

  1. YU:  ano  basuketto  o   motte,  (0.5) etoo, (0.5) sattetta no   ne,
          that basket      OBJ  carry         um          left     COP  ok
          he takes that basket and (0.5) um (0.5) rode off, OK,

→ 2. AZ:  hai
          yes
          [
  3. YU:  de   saroo to shitara, (continues speaking)
          And  leave about-to-do
          And as he is about to leave,
```

The transcript above is obviously quite simple. I have purposefully left out gaze and gesture as they are coded in very few papers[82], and they would make the transcript less approachable in that it would require an extra two lines of gloss. What the transcript does provide is a good idea of relative timing of the talk and the location of the recipient's response within the unfolding talk. AZ's response comes clearly after a pause which is also marked by a form of the copula, *no*, which typically signals the

[82] Szatrowski (2002) is notable for including detailed coding of gaze in the transcripts.

grammatical end of a sentence[83], and is followed by a particle, *ne*, which has been shown to have interactive and turn-related functions (Tanaka 2000). This transcript may tell us enough to make good inferences about when and how the participants are interacting.

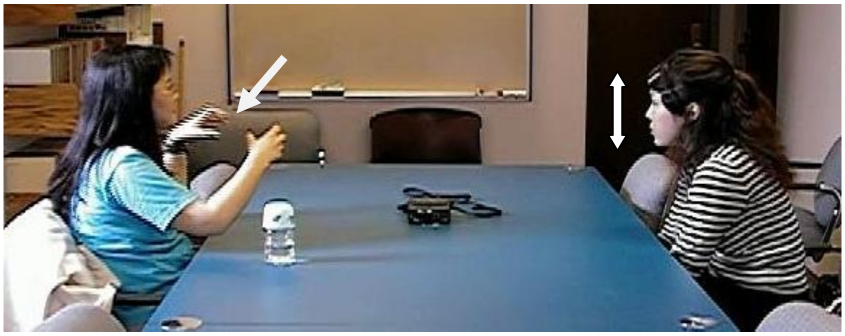
Even so, the video stills reveal a great deal more about the particpants' interactions. Yuri (YU) is on the left and Azusa (AZ) is on the right. In line 1, below, Yuri is looking away from Azusa, focusing on her own gesture. She stretches her arms out from her body as if placing the basket on the bicycle's front rack.



```
1. YU: ano   basuketto o
       That basket
```

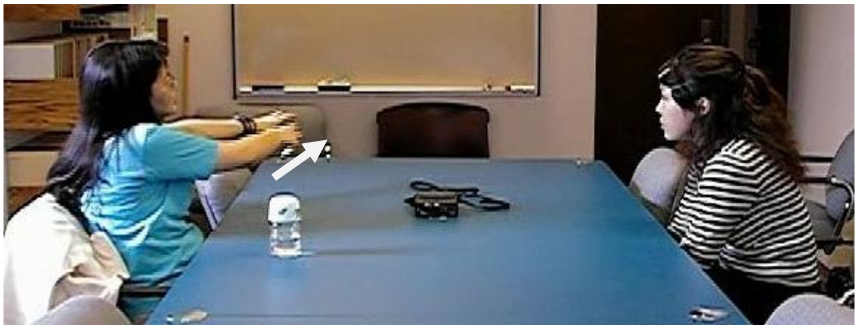
Next, Yuri brings her hands towards her body, acting out the word *motte* (line 2) which means "carry" or "take." *Motte* is produced in a conjunctive form (the so-called 'te-form'), so it is clear that there is more to follow. Azusa produces a series of very small nods (indicated with arrow). Yuri pauses for half a second, possibly considering how she should proceed.

[83] Of course grammatical 'sentences,' at least in the literary sense, do not always appear in talk. A speaker's utterance may consist of a single word (Schegloff, Sacks, and Jefferson 1979: 702) phrase (ibid), clause (ibid:703) or a series of chunks spread across several turns at talk that, if considered together, create a very sentence-like whole (Goodwin 1979:98). Nevertheless, the copula's association with 'the end' of whatever the speaker is talking about remains as a potential factor in recipient production of responses.



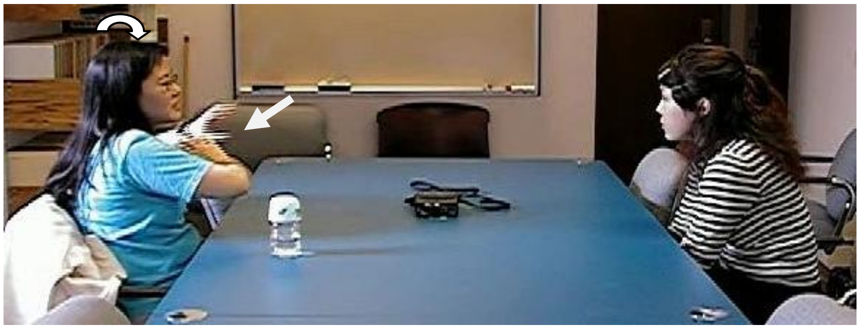
```
2. YU:  motte (0.5)
        took and
```

Next, she again stretches out her arms in a pushing motion and says, *etoo* “umm” (line 3). At this point it is obvious that she is struggling to find the word she wants, and it is unclear what this gesture indicates.



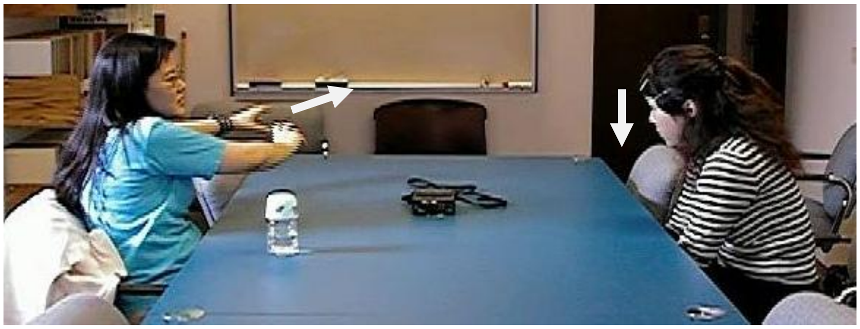
```
3. YU:  etoo (0.5)
        umm
```

During her pause, Yuri retracts her arms and gazes at Azusa’s face.



```
Yuri retracts her arms and gazes at Azusa.
```

As she moves her arms out once more, Yuri says, *sattetta no ne* “(he) left (and didn’t return)[84]” (line 4). As soon as Yuri completes the word *sattetta*, Azusa nods her head. After Yuri completes *no ne* “okay?” Azusa says *hai* “yes” (Line 5), and Yuri simultaneously begins Line 6.



```
    4. YU:  sattetta no   ne.
            left, OK.
                    [
       AZ:          N
 →  5. AZ:                   hai.
                             Yes.
                             [
    6. YU:                   de   saroo to shitara, (continues speaking)
                             and when (he) is about to leave,
```

The above stills demonstrate the value of video in several ways. First, the importance of Yuri’s gesture to the ongoing talk can be seen in the way both participants attend it with gaze. Yuri’s pause in line 3 represents a word search, though she is able to produce a gesture that indicates the meaning of her desired word, based first on the observable qualities of the gesture: Yuri is pushing away with her hands; and also on the fact that when she finally does produce the word, *sattetta* “left (and doesn’t return),” she repeats the same gesture again.  Yuri’s gaze shift may be to monitor Azusa’s understanding of the verb-gesture combination. Although the gesture along may have been understood within the context of the narrative as “go away,” the implication “and not return” likely would not have been clear.

[84] The use of this verb implies that the subject left and does not appear for the rest of the film.

Second, the video stills reveal how Azusa’s responses, a head nod along with the word *hai* in line 5 are produced at slightly different places but are no doubt connected. Azusa nods as soon as she hears Yuri say *sattetta*, but her verbal response is delayed a few tenths of a second until the grammatical end of Yuri’s utterance in line 4, possibly to avoid overlap. However, overlap occurs anyway in lines 5/6, likely because Azusa has already produced a head nod which Yuri interprets as a continuer. Overlap is common in transcripts presented in Maynard (1989), Yamada (1992), and Tanaka (2004), but it is possible that many instances of overlap can be explained by the timing of speaker gaze establishment as well as the sequential nature of recipients’ production of nonverbal versus verbal responses.

Third, the role of speaker gaze in recipients’ production of responses, as discussed in the quantitative section above, is hopefully made clearer. There, recipient responses were shown to be produced during units of speaker’s gaze, but not when the speaker was looking away. In the example above, Yuri establishes gaze at a specific point in her talk, namely after she has had some difficulty producing the word she wants to use. Her gaze change is quick and obvious, and Azusa responds, nonverbally, immediately after Yuri produces the word, *sattetta* “to go (and not return)”. Azusa is therefore responding to both the content of Yuri’s utterance and her gaze, which may represent a comprehension check following the successful completion of the word search. All of this could be explained through the use of a transcript alone, but the additional coding required would make it less approachable and more prone to error. Describing Yuri’s gesture in words would require a great deal of space on the page, but with the video still and an arrow or two, the scene is suddenly much easier to comprehend. In the next examples, where gaze and gesture play a greater role, a transcript alone will be shown to be inadequate for the analysis.

## 4.3 Recipient Responses and Speaker Pause

In this section I would like to approach the relationship of gaze to the occurrence of pause in a speaker's talk. 'Pause,' 'silence,' or 'zero uptake' can be interpreted as having meaning within the context of talk if it represents, for example, a recipient's refusal to provide feedback, but it can also indicate a word search, thought process, or even a distraction on the part of a speaker. Video data can obviously make the interpretation of pause much easier.

Some backchannel studies categorically define backchannels as a sort of non-intrusive feedback (e.g. Sugito 1987; Tanaka 2004) which "do not display disagreement or request more information[85]." Schegloff's (1982) paper describes "uh huh etc." [86] as "continuers" (81), though near the start of his paper he describes anecdotally the consequences of nonverbal feedback on the speaker's following talk. Specifically he mentions "the wrinkling of brows…a few smiles or chuckles or nods, or their absence" (2), which, in the context he discusses (a lecturer speaking to a class) might not violate the continuer function due to the situational roles of the lecturer, who by the nature of the situation controls the talk, and the recipients, though their actions could, as he says, affect the course of the speaker's talk. His actual analysis, unfortunately, does not approach such behaviour, so the question of whether his anecdotal examples function as continuers in conversation, where speakers and recipients are not bound by socially established roles, is left unanswered.  However, Goodwin (1980) demonstrated that

[85] The oft-described correlation between *aizuchi* and displays of harmony is less surprising when actions that may display disharmony are not accounted for.

[86] It is not clear exactly what expressions fall under "etc" in Schegloff (1982).

recipients' gaze (or rather, a lack of it) greatly affects the speaker's talk. This is also true for Japanese, as I intend to demonstrate, and appears to be a relatively intrusive form of feedback.

As stated earlier, speakers can pause during their talk for a variety of reasons. In the case of Japanese especially, speaker pause is sometimes attributed to a lack of uptake (in the form of a head nod or short utterance) by the recipient. Here I would like to demonstrate how analyses which limit coding to a restricted set of behaviours or those which do not utilize video can result in erroneous results.  In this short selection, the speaker, Akane (A) is describing the part of the Pear Film in which the main recurring character, a young boy, is about to take a single pear from a basket, only to change his mind and ride off on his bicycle with the whole basket. This transcript has been prepared to include vertical head nods (coded "N") but no other nonverbal actions, as appears to be common in the literature. The focus of this analysis is the one-second pause between Akane's utterance in Line 1 and Nobu's (N) response (vocal with a head nod) in Line 2.

### 4.3.1 A-N Transcript 1

**Data: [AK-NB GSC_0004.MPG]**

```
        1. A:  saisho wa   ikko dake totta n ya   kedo moo     kago  goto  motte
               first  TOP  one  only took    COP  but   already crate whole take-and
               first (he) only took one (pear) but then he took the whole crate and
                                               [
        2. N:                                  N

        3. A:  itchau n ya   n ka.
               go       COP   INT
               left

→       4.     (1.0)

        5. N:   n
                Uh huh
                [
                N
        6. A:  honde, sore  wa   jitensha ni tsunde, (continues narrative)
               then   that  TOP  bicycle  on load-and
               then he loaded that on the bicycle and,
```

The usual explanation for the one-second delay between Akane's utterance in line 3 and Nobu's response in line 5 is this: Akane will wait for Nobu to display understanding by means of a response[87] (head nod and verbal "*n*") before continuing on in Line 5. Nobu's response is delayed, causing Akane to pause. The speech environment for a recipient response is right: Akane's talk is grammatically finished, there is a clear low area of intonation at the end (underlined for visibility)[88], and there is a pause.[89] However, it is known that these criteria cannot predict recipient responses (Maynard 1989:175); they only point out the possibility of a response being produced there. I will demonstrate this in the next selection, which continues from the transcript above, in which Akane produces an utterance that ends with identical syntactic and prosodic features as the above, but does not include any uptake by Nobu or a pause.

In Line 7, below, Akane produces a nearly identical utterance ending,[90] *yan ka* "isn't it?" as she did in Line 3 (underlined in both). Despite the apparent lack of uptake, Akane continues speaking, with Nobu collaboratively completing the last part of Akane's utterance (Line 8). Obviously syntax and prosody alone cannot explain Nobu's choice to respond late (Line 5, above) or not respond at all (Line 8, below), nor can it explain Akane's decision to pause (above) or continue (below).

[87] Tanaka (2004) says that "a delayed *aizuchi*" may indicate apathy and subtly discourage the speaker (138).

[88] The duration of low pitch here is 130ms, more than enough to prompt a backchannel as per Ward and Tsukahara's (1999) findings.

[89] Pause is one of the main criteria for Maynard's (1989) Pause-Bounded Phrase Units (PPUs). In this case, the question here is, "why is the pause so long?"

[90] Not only are the utterances similar syntactically, the pitch curve (as measured using Praat) is nearly identical, exhibiting a 130 ms stretch of low pitch which could prompt a response, according to Ward and Tsukahara (1999). However a response is produced in only one of two cases here.

### 4.3.2 A-N Transcript 2

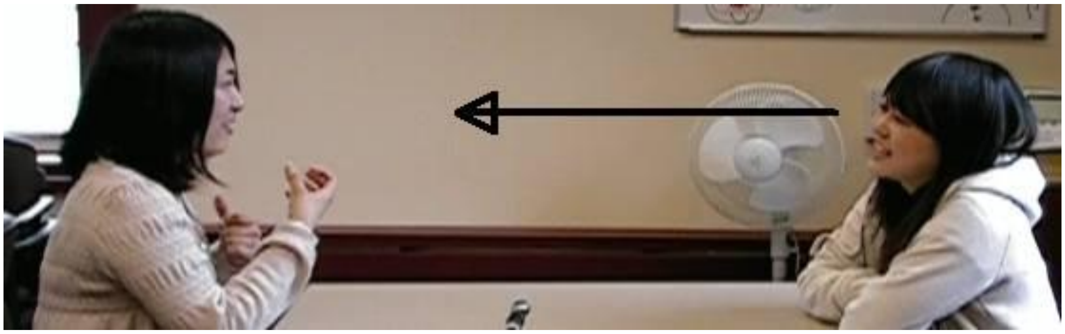
```
→       7. A:  mou      nigete  itchau  ya n ka.  Ossan    [ kizuite    nai  ya n ka.
               Already  run-and go       COP  INT. Old-man  notice  not  COP  INT
               then he runs away, right? The old man doesn’t notice, right?

        8. N:                                              [ kizuite nai.
                                                             Notice  not
                                                             doesn’t notice.
```

Widening the scope of analysis to include speaker and recipient gaze reveals a better explanation. Selection 6.1.3, below, is the same as 6.1.1, with gaze information added. After Akane says “*motte”* in Line 1, Nobu breaks mutual gaze by looking up and holding her gaze up until she responds with “*n*” and a head nod in Line 6.

### 4.3.3 A-N Transcript 1 + gaze information

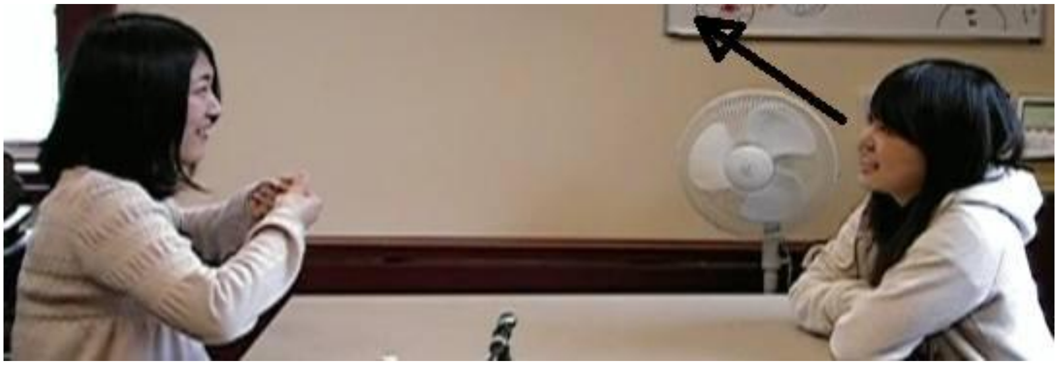
```
        Data: [AK-NF GSC_0004.MPG]

        1. A:  saisho wa   ikko dake totta n ya   kedo moo     kago  goto  motte
               first  TOP  one  only took   COP  but  already crate whole take-and
               first (he) only took one (pear) but then he took the whole crate and
                                                  [
        2. N:                                     N

        3. A:   itchau  n ya   n ka.
                  go       COP   INT
                went, OK?
               [
→       4. N:  <GAZE UP --------------------------------------------------->

        5.      (1.0 Pause)

→       6. N:                  n
                               uh huh
           N:   <GAZE RETURN>  N

        7. A:  Honde, sore  wa   jitensha ni tsunde, (continues narrative)
               Then   that  TOP  bike     on load-and
               Then  he loaded that on the bike and,
```

The following pictures illustrate this gaze change more clearly. Akane describes the boy taking one pear, and then as she says *kago goto* “the whole crate” (Line 1) she starts a gesture of scooping the basket of pears up and then holding it.



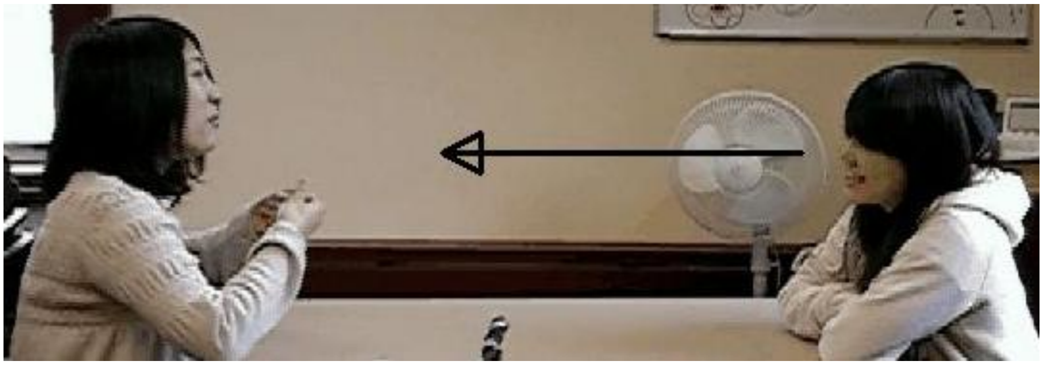
```
1. A:  saisho wa   ikko dake totta n ya     kedo moo     kago  goto  motte
       First he only took one (pear) but then he took the whole crate
                                            [
2. N:                                       N
```

After Akane says "*motte*" (end of Line 1), Nobu suddenly shifts her gaze up so that she's looking over Akane's head or at the ceiling (Line 4). Akane has been gazing at Nobu while speaking so Nobu's gaze shift must be obvious to her. When Akane comes to the grammatical end of her utterance (Line 3), she stops speaking, and her gesture remains frozen as if holding the basket.



```
3.  N: <GAZE UP ------------------------------------------->
        [
4.  A: itchau  n ya   n ka. (1.0)
       and went, ok?
```

After a pause of one second, Nobu brings her head down in a nod, says "*n,*" and resumes gazing at Akane (Line 5). Akane immediately resumes her narrative and her gesture (Line 6).



```
5. N: <GAZE RETURN> N  n.
                       uh huh
6. A:  Honde, sore wa jitensha ni tsunde,(continues narrative)
       Then, he loaded that on the bike and,
```

The pause highlighted above is clearly the result of Nobu’s gaze shift, to which Akane responds by pausing her talk at the end of her intonation unit (marked prosodically by falling pitch and a conjunctive verb form). In this way Nobu is able to temporarily halt Akane’s narrative. Once Nobu re-establishes mutual gaze and produces her verbal response, “n” and a head nod (Line 5), Akane continues. Gaze shift has not typically been considered in previous Japanese backchannel studies, but the above example suggests that recipient gaze may be connected to the production of continuers such head nods and short verbal responses (i.e. “backchannels”) that follow in sequence. Her responses therefore are not only in relation to Akane’s ongoing talk, but to her own previous action (breaking gaze) which caused Akane to pause. The pause is due to a nonverbal negotiation between Nobu, who breaks gaze, and Akane, who seems to interpret Nobu’s action as a sign of potential trouble.

It is likely then, that Japanese recipients may have a great deal of influence over a speaker’s talk through the use of gaze, just as English recipients do, as demonstrated by Goodwin (1981). In a similar vein, Ford, Fox, and Thompson (2002) demonstrate how a speaker faced with a non-gazing recipient at a possible completion point may produce an increment (and nonverbal gesture) in the hopes of attracting the gaze of that recipient or another recipient[91]. Although I believe the system at work is the same, as is the overall speech environment,[92] there is a fundamental difference in the example between Akane and Nobu above. In Ford, Fox, and Thompson's data, the recipients are either distracted by something or demonstrate their disinterest in the story by not gazing at the speaker at the speaker's possible completion point. In contrast, Akane and Nobu are already gazing at each other when Nobu suddenly breaks gaze in the midst of Akane's talk. Notably, Nobu breaks gaze to look upwards, to a place (the ceiling) where it is unlikely that something distracting could occur[93]. I will refer to this as the "thinking face," which is a special kind of gaze break that doesn't seem to coincide with speaker production of increments or other attempts to re-establish mutual gaze, and I will provide more examples from other participants in the next section. Now, I would like to look at an instance in which Akane produces an increment due to a lack of uptake from Nobu despite the achievement of mutual gaze, as a comparison to the findings of Ford, Fox, and Thompson (2002).

## 4.4 Recipient Responses and Speaker Repair (Increment)

A speaker obviously does not respond to all recipient actions with a pause, nor does a recipient have only gaze shift as a potential device. In the following example, I would like to demonstrate how the speaker reacts to visual cues when monitoring recipients during a point of mutual gaze. Specifically, the recipient makes a visually obvious change in her posture which the speaker notices and responds to upon

[91] Data is from English conversation.

[92] Technically speaking, the pause in 6.1.3 does not happen at a possible completion point, at least syntactically, but Maynard's (1989) demonstration of feedback at PPU boundaries and Iwasaki's (1997) analysis of Japanese speech production based on intonation units suggest that recipient feedback is highly likely at such a point, even if turn changes do not typically occur there.

[93] Nobu's gaze shift to the ceiling does not draw Akane's gaze as it might to an event within the horizontal plane, nor as it would to a gesture (Streeck 1993).

establishing mutual gaze. In the following excerpt A's narrative has reached the beginning of the final scene of the Pear Film, which occurs at the same place as the first scene.

Akane has just finished explaining the part where the boy gave the three other boys each a pear. But instead of continuing the narrative from there, she describes the overall path the first boy has taken until that point, which is shown in Lines 1-2: *honde, de sono otoko no ko wa sono ossan ga ita tokoro to, koo kita yan ka* "Then, that boy gets to where the old man was like this so it ended up like this." This is followed by a 0.5 s pause in Line 3, and then Akane produces the increment[94], "*ano ikisaki ga* "um, the destination" in Line 4, seemingly as a repair. Nobu responds, *n? n?* (Line 5) with rising intonation, demonstrating a lack of understanding or a request for more information, during which Akane expands the repair sequence by giving a condensed version of the entire story from Lines 6-12 and beyond, while Nobu makes regular feedback.

### 4.4.1 A-N Transcript

```
     AK-NB GSC_004.MPG (3:09-3:16)   N=head nod

     1. A:  honde, de   sono  otoko no   ko    wa   sono ossan ga   ita     tokoro
            Then,  and  that  male  GEN  child TOP  that uncle SUB  existed place
            Then, that boy gets to where the old man was like this

     2.     to, koo        kita kara koo        naru    ya n ka.
            and like-this  came so   like-this  become  cop INT
            so it ended up like this, right.

     3.     [0.5]

→    4.     ano  ikisaki     ga.
            that destination SUB
            Where he ends up.

     5. N:  n?  n?
            Huh? Huh?
                [
     6.         futsuu ni  koko ni  ki   ga  aru     yan
                As usual   here in  tree SUB exist   COP
                There's a tree like normal, right.
                                                   [
```

[94] Fox, Ford, and Thompson, 2002.

```
 7. N:                                            n
                                                  uh huh
    N:                                            N

 8. A:  kore ki   to shitara,
        This  tree as  if
        If this is the tree,
 9. N:                        n
                              uh huh
    N:                        N

10. A:  otoko no   ko    wa,  koo        kite, a, nashi aru   to   omotte,
        Male  GEN  child TOP  like-this  came    ah pear  exist QUOT  think
        The boy comes like this and thinks "ah, there are pears,"
                                              [                 [
11. N:                                        n                 n
                                              uh huh            uh huh
    N:                                        N                 N

12. A:  totte koo        itta n yan  ka
        take  like-this  went  COP   INT
        (he) takes one and goes off like this, right.
                                   [
13. N:                             n
                                   uh huh
    N:                             N
```

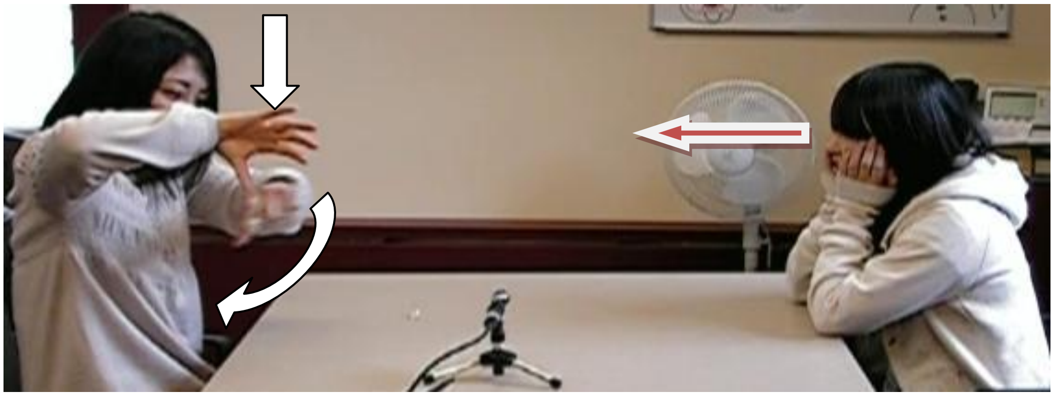
### 4.4.2 A-N Video stills

Akane's increment in line 4 is most likely related to the preceding pause. To demonstrate why Akane has decided to start a repair here, I will again provide a sequence of video stills. Note that the English gloss will be presented slightly differently to account more accurately for what is being said in each image. Just prior to Akane's utterance in line 1, below, Nobu changed her posture from her normal sitting position (depicted in 6.1.3) to what is depicted below: her head in her hands, elbows on the table. Nobu's posture shift appears to have gone unnoticed by Akane, who has been looking at her own gestures while talking. It is unclear if Nobu's posture represents some kind of embodied feedback in and of itself.[95]

Akane has not gazed at Nobu since Nobu changed her posture. In line 1, Akane starts talking about the relative positions of the boy and the man. As she says *sono ossan ga ita tokoro* "the place where that old man was" (line 1), her right hand, spread

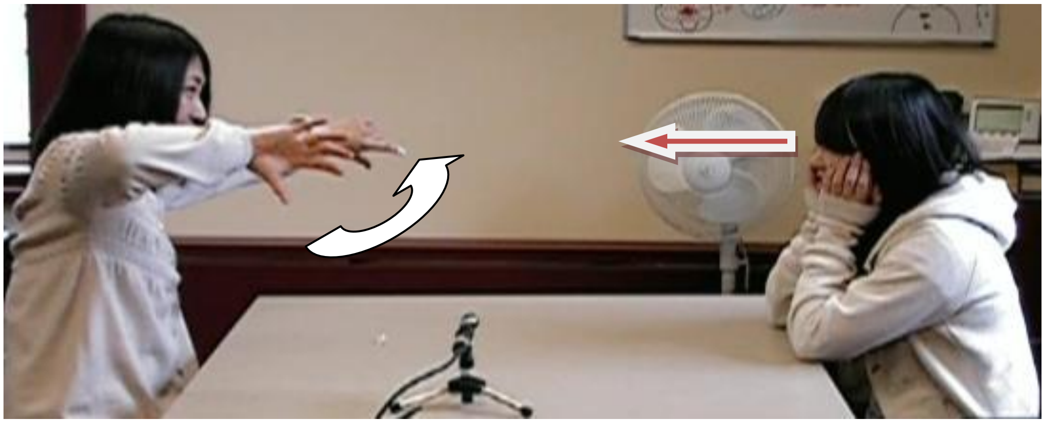
[95] I do not believe there is any difference between the use of this posture in English and Japanese.

wide, makes a downward motion as if placing the *ossan* in his proper place. Her left hand moves under and to the right to denote the starting place of the boy.



```
1. A: honde, de sono ossan ga ita tokoro
      then, where the old man is
```

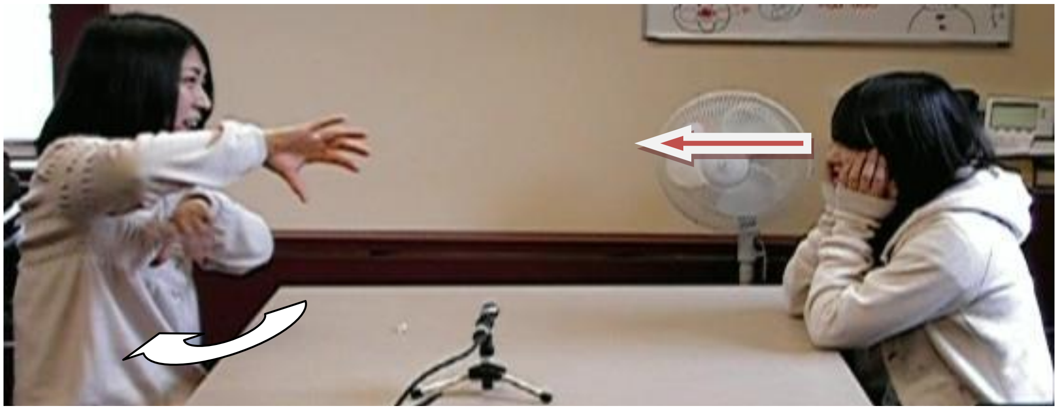
Next, she traces the path of the boy by swinging her left hand across and to the left. As her hand reaches the leftmost point of its arc, she looks at Nobu and nods when she says *yan ka* (line 1 a). As can be seen in the picture below, there is absolutely no response (other than continued gaze) from Nobu, either verbal or nonverbal; her expression[96] and posture remain exactly the same as before. Akane holds her hands at the end point of her gesture for 0.5 s. Mutual gaze continues.



```
2) A: to, koo kita kara koo naru yan ka, (0.5)
      he comes like this so it becomes like this
```

[96] It is possible that Nobu made an eye-area action, for example an eyebrow raise, but unfortunately her eyes are not visible to the camera.

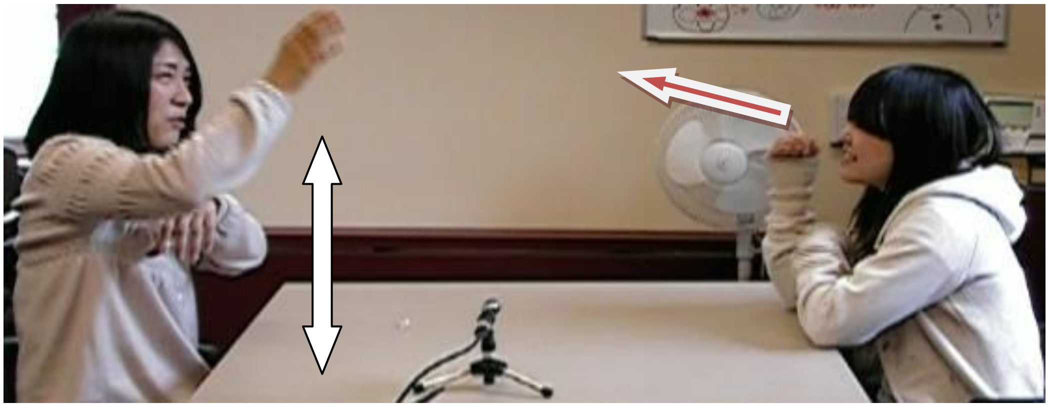
Continuing to look at Nobu, Akane very quickly moves her left hand back under her right hand and points down to denote the location she is trying to explain, and produces the increment, *ano ikisaki ga* “the destination.” In effect, Akane’s increment is a type of self-repair supported by the repetition and slight modification of her non-verbal gesture, probably in response to Nobu’s lack of uptake.



```
3) A: ano ikisaki ga
      where he ends up
```

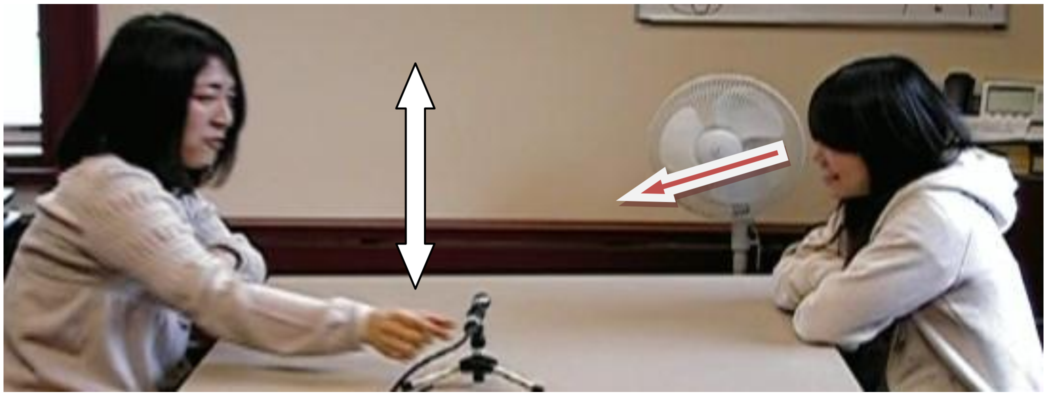
Immediately after Akane’s increment in Line 3, *ano ikisaki ga*, Nobu begins moving her hands away from her face, and her head tilts upward rapidly as she says “*n? n?*” with rising pitch (Line 4), signalling interest and possibly confusion about Akane’s description of events. Her hands begin to move down from her face, and her head moves up, while her body remains slightly tilted forward.

While Nobu is shifting her posture, Akane has already begun another sequence in which she elaborates on her previous repair by explaining, in detail, the relative locations of all the characters and the tree, which acts as the point of action for both the start and end of the Pear Film and ties the actions of the characters together. It is likely that Akane has noticed Nobu’s posture shift at this point, though she is not looking directly at her.



```
4) N: n? n?
      huh? huh?
          [
5) A:     futsuu ni ki ga aru  yan
          there's a tree like normal right
                                 [
```

It appears that Akane correctly predicted the trouble source, because N returns to her ‘baseline’ posture and begins verbal and nonverbal responses again, the first simultaneously with Akane’s “*yan*” in Line 5.



```
                [
5)   A:      futsuu ni ki ga aru yan[97]
                                   [
6)   N:                            n
     N:                            N
7)   A: kore ki toshitara
8)   N:                   n
     N:                   N
```

[97] This line is repeated from Line d in the previous picture.

Just like in Fox, Ford, and Thompson's (2002) examples from English conversation, Akane's increment in line 3, above, achieves a recipient response from Nobu. Nobu's response is not just a simple continuer, but a combination of re-posturing, head movement, and an inquisitive verbal utterance, *n? n?* which signals a return to active listenership and a request for further explanation. Considering the scope of typical Japanese backchannel studies, most of Nobu's nonverbal behaviour would not have been coded despite its importance to the continuing talk.

One concern that arises from this example is that of classification of repair. Relying on the transcript from 6.4.1 without looking further at the video data would result in classifying Akane's repair as self-initiated self-repair. But the video stills show that she is obviously affected by N's non-response (and possibly her posture), making the argument that this repair is other-initiated a good one. Ultimately it is Akane herself who determines when and what she needs to repair, but N is by no means a passive listener, despite her silence. Her lack of response (whether for a purpose or otherwise) is responsible, at least in part, for Akane's repair initiation, and once the repair sequence begins, her resumption of verbal responses and head nodding suggest she is satisfied with the repair attempt. Without video evidence, this sequence would be impossible to accurately categorize. But even after looking at the video evidence it is difficult to determine what move or combination of moves led to Akane's repair initiation: was it Nobu's silence, her posture, her unchanging facial expression, the lack of a head nod, or (the most likely scenario) some combination of these? A rigorous CA-based approach to better determine the sequential relationship between speaker gaze (and head nods) and recipient responses in terms of adjacency pairs[98], for example, would be useful to better understand what exactly is going on here.

So far I have looked at how recipients' actions can affect the speaker's ongoing talk through the use of gaze, bodily movements, and verbal responses. But recipients can utilize these same devices to actively take a turn at talk.

## 4.5 Recipient Gaze Change as Turn-taking Device

So far, I have demonstrated how gaze is used by a speaker to create the opening for a recipient response (section 6.2); following that I showed how recipients can, instead of producing a continuer-style response, may break gaze (section 6.3.3) or shift their posture (6.4.2), which leads the speaker to either pause or perform a repair. Below, I will show how a recipient can use gaze shift via the 'thinking face' to actively attempt to take a turn at talk. Again I will start with a transcript which includes only head nods as nonverbal behaviour, and then follow up with more evidence from video stills.

In the sequence below, Aiko explains the setting of the Pear video to Yoshi. Near a possible grammatical completion point of her utterance, which continues through lines 1-4, Yoshi interjects with a loud "aa" (line 5) before asking a question. The speech overlap here is the main point of interest. Why does Yoshi jump in here? Maynard (1990:410) noticed that Japanese interlocutor responses tend to overlap more often than American English speakers', especially near the end of grammatically complete utterances. In the transcript below, Yoshi produces one head nod, coded "(HN)," shortly into Aiko's utterance in line 1. He produces no more head nods or backchannels until he takes the turn in line 5. It is therefore possible that Aiko interprets his lack of uptake as a trouble source and yields the turn, as her voice decreases in loudness as she says "*desu*"

[98] An adjacency pair in CA is a two-part sequence beginning with an utterance by one speaker and finishing with a response by another, found in, for example, greetings.

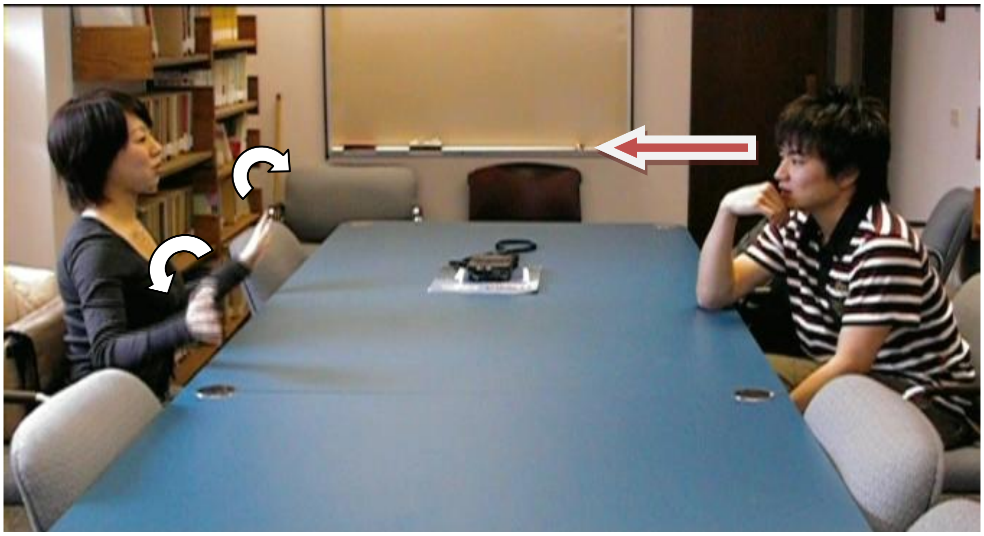
(line 4). However, Yoshi has already started speaking at this point, so Aiko’s prosodic change is not likely a sign that she is yielding the turn. Yoshi could be predicting the end of her utterance based on the projection of grammatical completion of Aiko’s utterance, signified by the noun *tokoro* “place” in line 4, although this is unlikely as the same noun appears in line 3, followed by a pause and then an increment.

### 4.5.1 A-YM Transcript

```
   Data: A-YM

  1. A:  kemono michi mitai-na   michi o    hito    ga   riyoo shite-iru
         Animal path  looks-like path  OBJ  person  SUB  use   do-ing
         there’s a path that looks like an animal path people are using
                                 [
  2. Y:                          N

  3. A:  tte   iu   yoo-na  tokoro  .. sugoku  soo- soodai-na, shizen  no,
         QUOT  say  like    place      really       grand      natural GEN
         a place like that .. a really grand, natural

  4.     tokoro desu.
         place  COP
         place.
               [
→ 5. Y:        aa e  mawari  ni   ie    toka    wa   aru   n desu ka.
               Ah Eh around  LOC  house example TOP  exist   COP  INT
               Ah eh are there like houses and things around there?

  6. A:  ie     mo     hitotsu mo   ikke       mo   miemasen.
         House  even   one     even one-house  even can’t-see
         There isn’t even a single house.
```

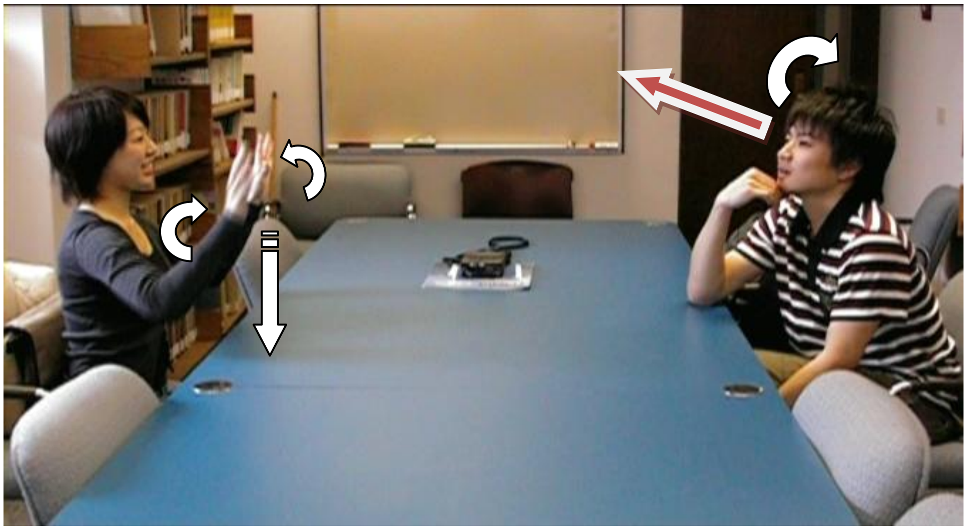
But if we consider the original data in its entirety, things are not so simple. By considering all the body moves that occur during this brief interchange, and not only head nods, a much more complex yet realistic explanation is revealed.

**Last half of line 3 (below):** Aiko, on the left, is describing the setting of the “Pear film.” She has brought her gaze to Yoshi, on the right, who is already looking at her. Aiko's hands are moving up and out as she describes what a grand vista the film showed. Yoshi's posture, expression, and body position are stable, and he gazes directly at Aiko, displaying his active listenership.



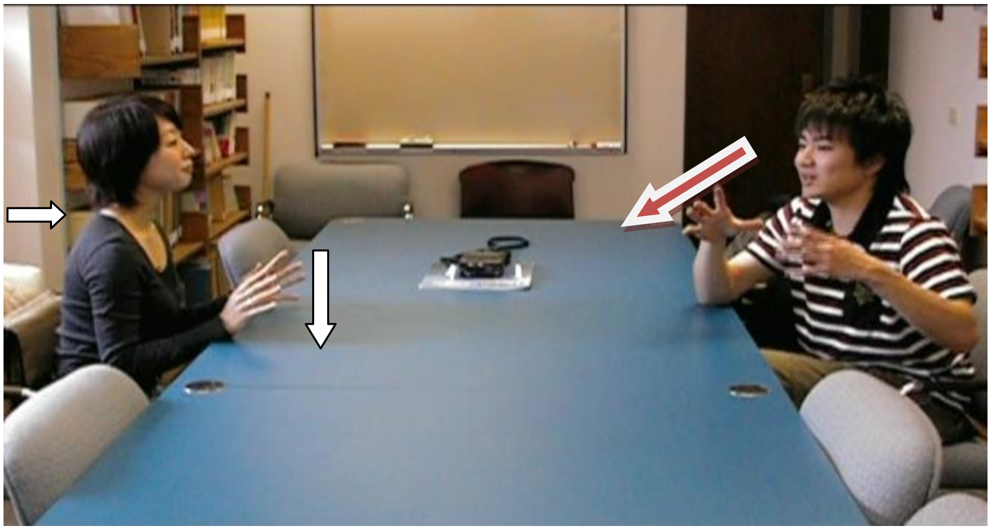
```
3. Aiko: (previous part omitted) sugoku soo-
                                 really gra-
```

**Line 3, continued (below):** As Aiko says “*soodai*,”Yoshi suddenly withdraws his gaze by jerking his head back and looks up as if in thought, and keeps this position. Aiko brings her utterance to a grammatical end. Aiko gestures by moving her hands back and forth in an arc in front of her chest several times. Each time she repeats the gesture it gets smaller and her hands move closer to the table.

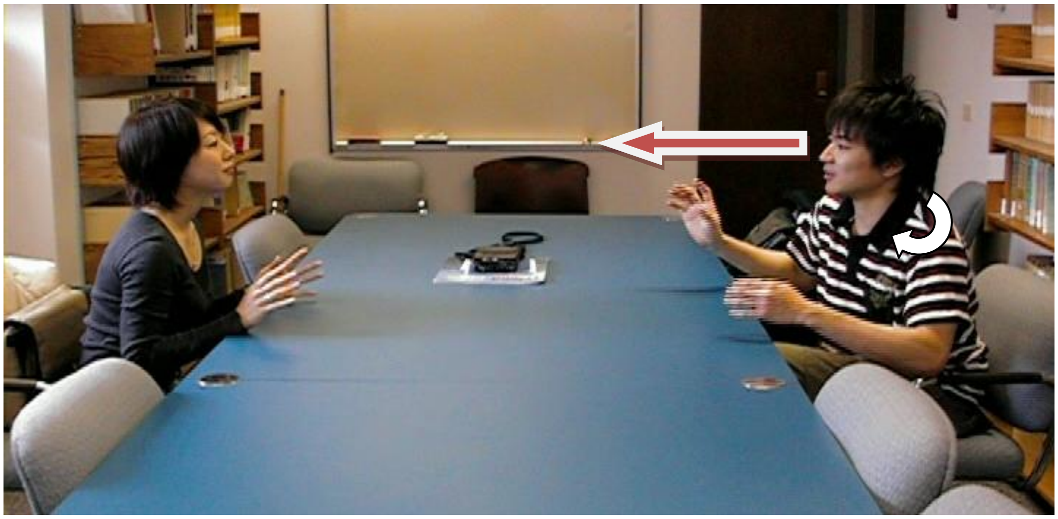


```
3.   Aiko:   (previous part omitted) soodai na, shizen no
                                     grand, natural
                                     [
     Yoshi:                          <Head jerks back>
```

**Line 3, continued (image below):** As Aiko says *tokoro*, she ends her gesture by placing the heel of her hands on the table, her palms facing up and her fingers still spread. She also changes her posture by leaning forward, possibly signaling that she is ready to listen. Since Yoshi is looking up it is not clear exactly what combination of Aiko's moves he uses to determine that it is his turn to talk, but he beings speaking exactly when Aiko’s hands stop moving, before Aiko completes her utterance grammatically. Still looking up, he begins by saying “*aaa*,” quite loudly, with which he sounds to be demonstrating his understanding of Aiko’s just-finished talk, or alignment to it. He then lowers his head asks her a question (Line 2), but faces her only once his utterance has been well established (end of line 5, bottom picture, below).



```
4. Aiko:   tokoro  desu.
           place.
                   [
5. Yoshi:          aaa  e mawari  ni  ie  toka wa
                   ah eh are there like houses
```



```
5. Yoshi:  aru n desu ka?
           around there?
```

The video data shows that the participants have a number of tools at their disposal to coordinate a turn change. Foremost is gaze; when Yoshi retracts his gaze and puts on a 'thinking face,' Aiko reacts by ending her turn. The fact that she is finished speaking in line 4 is not clearly projected by prosody or syntax, but her nonverbal moves, namely her posture shift and gesture stop are very salient. Note that Yoshi does not meet Aiko's gaze again until after he begins his utterance, which may be a tactic by which he can retain the turn until he is ready to give it up.

It is not likely that any of these body moves operates completely independently as they are produced sequentially with significant overlap. More likely the participants make a combination of moves and a portion of them are noticed and responded to[99]. In the above sequence, Yoshi is not simply avoiding gaze with his eyes; he visibly jerks his head up as well. Aiko's interpretation of Yoshi's moves is likewise probably based on a variety of factors as well, including his timing and the current content of talk. Nonverbal

[99] Hayashi (2003:167) similarly describes how speakers simultaneously produce more than one action: "… the speaker who engages in a word search deploys multiple practices in different semiotic modalities…" which "…provide enhanced projective sources for recipients to anticipate and produce the target of the ongoing word search."

moves and reactions to them are also obviously affected by the participants' roles. Take gaze, for example: a speaker makes and breaks gaze many times during talk without any interactional repercussions (as long as a recipient shows attentiveness), but when a recipient breaks gaze, it is attended to very quickly.

In the literature, backchannels are considered to be inherently *responsive*, for example as 'continuers and assessments[100].' Yet in the above example it is clear that Yoshi is using bodily actions and response-like behaviour not only as continuers or assessments but as actions that Aiko must also interpret and respond to; in this case she responds by allowing Yoshi a turn at talk. Note as well that although Yoshi's speech behaviours occur within a turn-based framework (albeit with slight overlap), his bodily actions are not constrained by this framework: he can make them at any point during Aiko's turn. But this is not to say they are independent of the turn system, since their interpretation depends on the context of the speaker's talk at the moment they are made.

I followed up on these observations by examining three videos of natural conversation and found that the same procedure is used, and in fact is used much more often (in terms of frequency) in those videos than in my narrative corpus. I attribute this to the fact that narratives are, by design, not as conducive to heavy recipient interaction and turn change as would be conversation, which involves a great deal of shared knowledge and opinion, and plenty of turn changes.

Although the recipients in the narrative data were seen making the 'thinking face' very regularly during the speakers' talk, it did not always lead to a change in the speaker's talk or a turn change. This is because recipients could 'cancel' their 'thinking

[100] Schegloff 1982.

faces' by producing a series of head nods and resuming gaze at the speaker, similar to how Goodwin's (1981) English participants performed mid-turn gaze breaks while continuing to provide feedback, demonstrating their continued listenership (106).

In this section I showed how recipients can negotiate a turn change by breaking mutual gaze with the speaker and jumping in at an opportune time using the 'thinking face.' Of course, the 'thinking face' is just one of many nonverbal negotiation behaviours I observed in the corpus. In the next section I would like to introduce very briefly another overlooked head movement in order to suggest that recipients have a wide variety of identifiable nonverbal tools at their disposal and that language investigators have much to identify and explain.

## 4.6 A previously overlooked response: "Head Twist"

Maynard (1989) examined only vertical head nods for her quantitative study, but she did recognize a variety of other head movements that occurred in the data. Likewise Ynvge (1970) and Schegloff (1982) list a variety of behaviours, from posture change to facial expression, which could have significant outcomes for the talk-in-progress[101]. Unfortunately for the Japanese case I have found no backchannel studies that go beyond vertical head nodding[102], and for that reason I would like to take a brief look at one type of recipient head movement that has been overlooked. In this section I will describe the use of a "head twist" motion in which the recipient cocks the head to one side or another, possibly to display a lack of understanding or a request for more information. In this data set there are only two examples of the head twist, performed

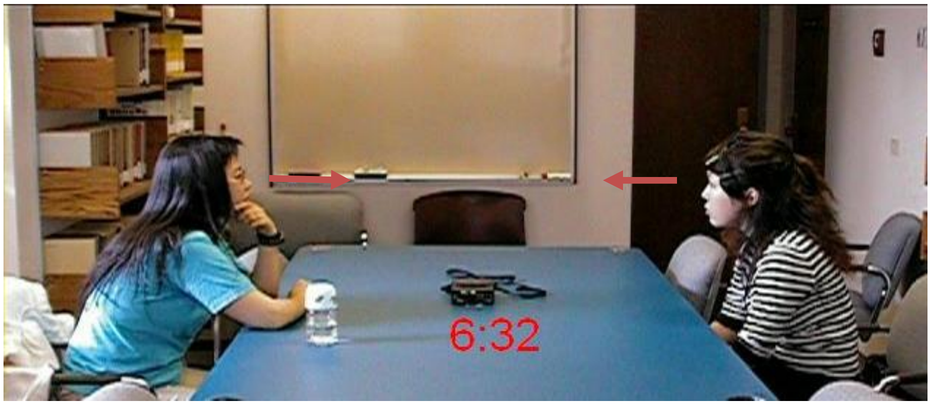
[101] Although Yngve (1970) and Schegloff (1982) both list several behaviours as potential candidates for analysis, most of the candidates were not actually analyzed.

[102] A major exception is Hayashi (2003) who performed a fully multimodal analysis on some instances of joint utterance construction. Backchannels and *aizuchi* were not the focus of the study, however.

by a single female recipient. No firm conclusions can or should be drawn from this data; the purpose of this brief analysis is to suggest the importance and complexity of other less frequent head movements and why we should make them a part of our analyses.

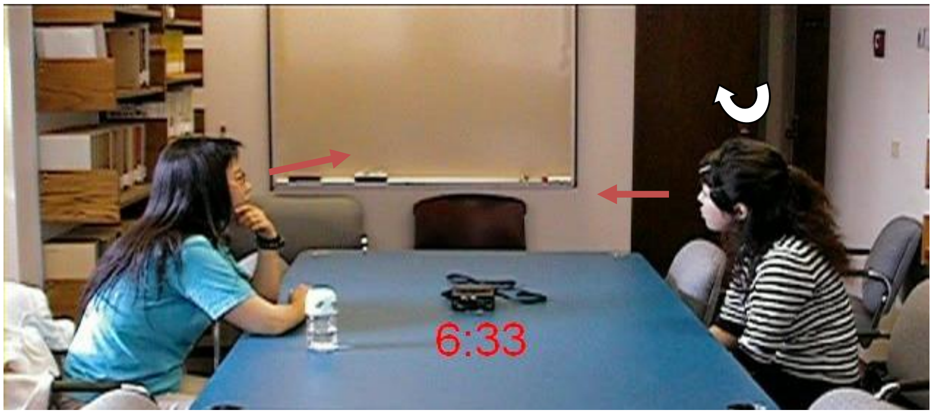
### 4.6.1 YU-AZ Video Stills

On Line 1, below, Yuri (left) is trying to describe her impressions of the old man who appears in the Pear Film. She asks Azusa if she understands the English term, "migrant worker."



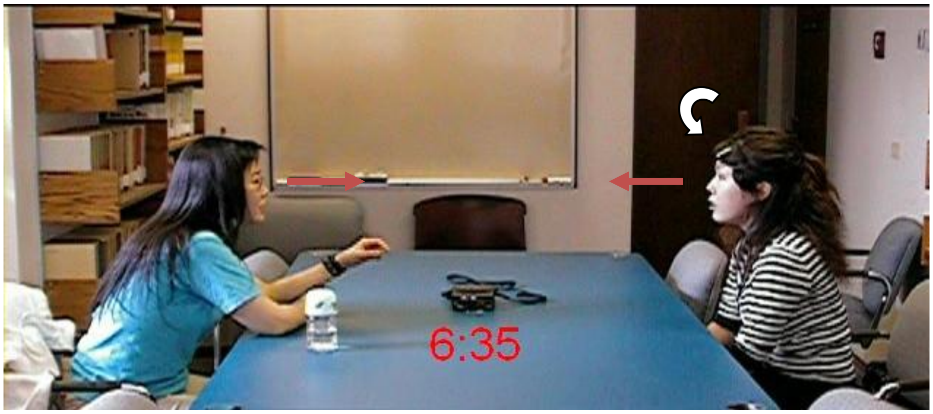
```
1. Yuri: Unto, 'migrant worker' tte wakaru?
         Um,do you understand 'migrant worker?'
```

Azusa responds immediately by performing a head twist (Line 2, below), demonstrating that she doesn't understand that word. Yuri continues by saying she doesn't know the Japanese term (Line 3). However, it is clear that she is performing a word search by her posture (body leaning forward, hand cupping chin), her statement, *nihongo wakannai n dakedo* "I don't know in Japanese but" and her gaze, which briefly shifts to the left (not visible in the picture). After Yuri's explanation in Line 3, Azusa produces a head nod in Line 4, with her head still cocked to one side.



```
2. Azusa:  <HEAD TWIST-------------------------
3. Yuri:   Nihongo wakannai n dakedo,
         I don't know in Japanese but,
4. Azusa:  ----------------------------> <HEAD NOD X 3>
```

In Line 5, Yuri snaps her gaze back to Azusa and makes thrusting hand motions toward her as she says the correct Japanese term, *kisetsu roodoosha* “seasonal worker.” In Line 6 Azusa responds by simultaneously saying *“aa!”* and moving her head back to the ‘baseline’ position. She then nods twice in sequence and says, “*hai*” ‘yes.”  Yuri then continues where she had left off before the word search.



```
5. Yuri:   ano, kisetsu roodoosha?
           Um, kisetsu roodoosha?

6. Azusa:  Aa!                Hai.
           <HEAD STRAIGHT> <HEAD NOD X 3>
           Ah!                Yes.

7. Azusa:  De, koo kite iru hito tachi .. kana? .. tte omotta.
           Well, they were like these people (in the film).. maybe? ..  I thought.
```

When the recipient, Azusa, produces and maintains her “head twist,” she is responding with a nonverbal action which may demonstrate alignment: that she is also trying to think of the word. It could mean she doesn’t know. What is clear is that she holds this position until Yuri has successfully completed her word search, physically and visually demarcating a “branching off” from the main topic into a negotiation space for Yuri’s word search. Azusa’s responses while performing the head twist are bound to the negotiation sequence, and do not refer to the main line of talk, i.e. they act as continuers for a ‘subspace’[103] negotiation, not for the main narrative. When Yuri provides the correct word, Azusa displays acceptance or understanding by moving her head back to the normal vertical position and then producing a head nod in that position, signalling a return to the main topic. In a way, Azusa’s change in head position could be considered a nonverbal style shift which creates or maintains a subspace in which a line of talk parallel to the main topic can be undertaken.

Although the ‘head twist’ may be relatively uncommon, I believe it is imperative for investigators to be able to account for all kinds of recipient behaviour. Unfortunately the methodologies that have dominated investigations into Japanese backchannels/*aizuchi* have, by design, overlooked many of these arguably less common but no less important behaviours. The terms ‘backchannels[104]’ and ‘*aizuchi’* of course have certain behaviours associated with them due to past studies and casual notions of

[103] Subspace is a term introduced by Ikuta (2008). In her paper, subspace is a duration of talk demarcated from the main talk through speech style shift. A full discussion of style shift is beyond the bounds of this thesis, but other than Yuri’s production in line 5 of the bare noun *kisetsu roodoosha* “seasonal worker” there appears to be no major speech style shift going on. What I am suggesting is that the participants can use nonverbal tools to branch off from the main topic of talk resulting in a similar outcome to what Ikuta (2008) observes in the use of speech style shifting.

[104] Yngve (1970) mentioned a very wide range of behaviours under the term ‘backchannel’ including head movements, but unfortunately they do not seem to have been investigated further in following studies on backchannels.

what they define; the word *aizuchi* especially has some socio-historical implications and was originally a colloquial term, not a technical one. For these reasons, behaviours such as the ‘thinking face’ or the ‘head twist’ are not generally included within their definitions. But as I have shown, both of these terms are poorly defined -- not a productive feature for research-related jargon – resulting in the vocabulary itself limiting investigators’ ability to approach newly identified behaviours or to modify the existing jargon to cope with new findings.

## 4.7 Summary of Qualitative Results

In the qualitative analysis I took a fully multimodal approach in which every potential interactive device: verbal, nonverbal, or even visual, was carefully accounted for. I selected a single salient recipient behaviour for deep analysis: the ‘thinking face,” and determined how it played out sequentially in several examples. I found that a single behaviour can result in multiple outcomes, ranging from a pause in the speaker’s talk to a turn change, depending on how it is deployed with other devices available to the recipient. I then took a brief look at another interesting head movement, the “head twist” to suggest that there are many other overlooked nonverbal actions that probably represent a wide variety of functional resources for participants.[105]

Avoiding speaker gaze or saliently breaking gaze with the speaker was found to be the main feature of the ‘thinking face;’ it is an action/request that is noticed immediately if mutual gaze has already been established or the next time the speaker attempts to achieve mutual gaze with the recipient. The ‘head twist’ was shown to be employed, in the single case that was looked at, as an independent response to a speaker’s word search initiation during a state of mutual gaze; no corresponding verbal

[105] Some examples include facial expressions such as frowning, pursed lips, eye rolling, and smiling; horizontal head shakes; and attention-getting arm movements and hand-waves.

component was necessary. All possible functions/outcomes of the ‘head twist’ have not yet been determined, yet it is obviously a device that is used as an indicator of listenership at the very least, but also possibly demonstrates alignment with the speaker.

# 5 Conclusions

This study took a both a quantitative and multimodal approach to further investigate some behaviours that have never been considered to be 'backchannels' or *aizuchi* in previous studies, despite the similarities of the environment in which they are produced, or the fact that they often occur simultaneously with those behaviours. The first part of the quantitative study was limited to certain behaviours that have been previously established as being quantifiable, specifically short response utterances and vertical head nods; additionally I attempted to expand on previous studies by also coding speaker head nods. In this section I found that a large number of recipient responses are preceded by speaker head nods, demonstrating that recipient responses may be more closely associated with speaker behaviour than was previously realized. The second part of the quantitative study expanded the focus to include gaze units, which, to my knowledge, have not been quantified previously. The state of mutual gaze between speaker and recipient was determined to be the main environment for recipient production of responses, as very few recipient responses were made in the absence of gaze speakers' production of head nods indicated where the majority of recipient responses occurred[106].

The qualitative multimodal section took a closer look at a specific example of recipient response behaviour that has not been examined before, namely the 'thinking face' in which recipients break or avoid mutual gaze by looking up. Another previously overlooked head movement, the 'head twist,' was also briefly described and was shown to have a very different outcome than the 'thinking face,' suggesting that a wide range of functional behaviours have yet to be identified and analyzed.

[106] Further investigation on gaze state and speaker production of head nods is necessary.

The quantitative and qualitative sections turned out to be complimentary. Initial qualitative observations of gaze behaviour in the corpus led me to design a gaze-based database, and the quantitative findings confirmed that gaze is indeed an important factor in recipients' production of responses.  The full results of the quantitative section led to some very interesting preliminary findings on the relationship between speaker gaze, head nods, and recipient responses. It also highlighted the degree of individual difference involved in recipient responses. Most importantly, it brought up new questions and led to the close qualitative analysis of behaviours that could not be quantitatively analyzed, culminating in the identification of a recurring recipient response behaviour, the 'thinking face,' which was found to be a highly interactive device that, depending on its deployment, has great effects on the ongoing talk, namely pause, repair, and even turn change.  On their own, both approaches are limited: the quantitative approach allows for the quick analysis of a relatively large amount of data, but requires identifiable instances of pattern-based behaviour for coding. On the other hand, the qualitative approach allows for detailed analysis but only on a small scale. Together, both qualitative and multimodal approaches lead to a valuable cycle of investigation and discovery.

For both analyses, video and high-quality audio data was favoured over the use of transcribed data in order to approach it in a way closer to how the participants experienced it. The quantitative study resulted in the observation that recipient responses are highly correlated with mutual gaze between speakers and recipients. Based on this observation, the qualitative portion of the study showed that a variety of recipient responses were being performed in coordinated manners in the same speech environment underscoring the importance of taking a multimodal approach to language and interaction.

## 5.1 Backchannels? *Aizuchi*?

The quantitative analysis demonstrated that there is great individual variation in the frequency, type, and combination of so-called backchannel responses or *aizuchi*. The most revealing finding was the relationship of recipient responses to speaker gaze establishment, suggesting that response production is highly contingent on a combination of speaker gaze, head nods, and talk rather than on talk alone.

In the qualitative analysis, I looked at a type of upward head movement or 'thinking face' that has not been considered in Japanese backchannel studies, despite its obvious use as a responding device in the same environment as head nods and short utterances such as *un* or *hai* "yes." Recipients' 'thinking faces' had obviously different effects on speaker talk: speaker pause, speaker increment/repair, and turn change. Although the environment in which the 'thinking face' is initiated can be anywhere within the speaker's turn unit, the effects of the 'thinking face' occur during speakers' attempts at gaze establishment with the recipient, i.e. the same environment as the more typically considered 'backchannels' or 'continuers.' Schegloff (1982) claimed that continuers are part of the system of repair (i.e. they pass up opportunity for repair); I believe that the 'thinking face' goes beyond indicates some kind of repair request or 'assessment' of the speaker's talk[107]. Rather than 'backchannels,' 'continuers' or other functional descriptors, perhaps these recipient behaviours would be better described or grouped based on their production environment within the turn-based system of talk.

[107] As described by Schegloff (1982:85).

I hope this thesis has demonstrated that the terms ‘backchannel’ and *aizuchi* are first, insufficient, and second, problematic for the purpose of linguistic analysis. The range of behaviours described in previous studies does not account for the complex multimodal interchanges of which they are a part. Nor are these behaviours ‘back’ in any way; they are produced and attended to as part of the ongoing talk.  I cannot yet suggest alternate terminology, but I would suggest that these behaviours are better considered as complex interactional devices that represent points of negotiation between speakers and recipients in ways that go beyond a ‘continuer’ or simple response function. Taking a blanket approach by grouping them in terms of quantity or visible qualities misses out on the fact that each and every one is contingent on what comes before, and consequential for what comes after. They are neither ‘back’ nor do they represent a separate ‘channel’ of communication; they are intertwined with the ongoing talk and other accompanying nonverbal actions.

## 5.2 Gaze

I hopefully have demonstrated that gaze is a very important factor in Japanese interaction. Points of mutual gaze establishment are highly interactive environments that need to be explored further. Likewise, the production of a variety of nonverbal actions, including head movement and gesture, by both speakers and recipients suggest that nonverbal or visual elements of talk in Japanese are important to a greater degree than previous studies have suggested.

Speaker gaze was established as the major determiner for most of the recipient responses in the data. Speakers’ active attempts to establish mutual gaze with the recipients created the environment for most recipient responses. When recipients chose not to meet gaze at points established by the speakers, the speakers responded immediately by pausing, performing a repair, or by allowing the recipients to take a turn at talk. Speaker gaze in the narrative data was found to be a major organizational resource for both speakers and recipients. The same kinds of gaze-related speaker/recipient interactions were observed in the conversational data, but because turn-changes occur so often in conversation, gaze is more difficult to code than it is in narrative talk.[108]

Although speaker gaze was found to be an organizing device, it did not follow that speakers were in total control of the direction of ongoing talk. Recipients used gaze-avoidance through the use of the ‘thinking face’ to visually demonstrate some kind of trouble which the speaker attended to via pause, a repair, or by allowing the recipient a turn at talk. The speakers had to determine what exactly the trouble source was and then adjust the talk-in-progress to accommodate for it, sometimes giving up the turn at talk entirely.

## 5.3 Baseline Behaviour and Deviations from the Baseline

Baseline behaviour was highly evident by the fact that speakers and recipients tend to take a more or less static posture, and return to that posture quickly after performing an action that requires them to deviate from it. Posture shifts, gaze breaks or large, noticeable movements by recipients that did not relate to the talk were performed while making repetitive head nods, which demonstrate a continuer function and also a display of continuing listenership. In other cases they were performed unnaturally slowly so as not to attract attention. In one instance (not shown), a recipient moved his keys and wallet from one pocket to another, requiring him to look away from the speaker, move his chair back, and make large adjustments

[108] I imagine that a gaze-based database would be possible for conversation as well, but it would be more complex since both participants switch roles so often.

to his posture. He did this very slowly, and as quietly as possible; his nylon jacket made this difficult. While he moved, he performed regular and continuous head nodding as if to tell the speaker “don’t mind me, keep talking.” In this case, the recipient used constant head nodding as a way to “cover” non-relevant actions that could attract attention and/or be interpreted as some other kind of feedback, while also displaying continuing listenership. A similar case in English is described in Goodwin (1981:87-88).

The careful observation of deviations from established baseline behaviour could possibly prove fruitful in the coding of multimodal data because it helps establish the interactional behaviours the participants consider to be relevant to the talk. The existing concept of marked/unmarked language use could possibly be applied to what I call baseline behaviour, although there are some important implications that make the application of the marked/unmarked dichotomy less desirable in this case. Determining what constitutes baseline or unmarked behaviour requires careful observation of each individual’s actions and will likely differ from situation to situation, due to the environment, topic of talk, conversation partner, and even individual mood. In other words, every interactive sequence has its own flow which must be determined. Thus what is ‘marked’ behaviour in one interaction may be ‘unmarked’ in another, which seems to go against the purpose of such categorization.

Identifying baseline behaviour involves, first of all, ‘getting to know the participants’ through close observation of the data. For example, participants in this corpus tended to adopt the same posture throughout the exercise; a major shift in posture was a clue that something was going on. Looking more closely at one example, I discovered that a recipient’s posture shift was accompanied by an upward head movement and upward gaze that began during the speaker’s ongoing talk and continued even when the speaker attempted to achieve mutual gaze. I coded this as a possible response in the database although at the time it didn't seem to correlate with anything known. Later on, after the quantitative results showed a relationship between speaker gaze and recipient response, this case became one of many examples of the 'thinking face' that were identified precisely because they were produced outside of the 'normal' environment for recipient responses, namely speaker gaze.

## 5.4 Issues: video & technology

I hope that my thesis demonstrates the potential benefits of video to linguistic analysis as well as the necessity to avoid the pitfalls of 'premature coding' (Gardner 2001). I based my analysis entirely on the original video data, but this also turned out to be very problematic. There are very few applications that can do everything a language analyst wants. Those that exist are very complex. I opted to use a handful of free programs which turned out to be very good in some ways but limiting in others. I could not find a program that would allow me to easily add annotations (specifically graphics) directly to the video[109].

## 5.5 Issues: transcription

Discussion regarding issues with transcripts is not new. Ochs (1979) discusses the problem of "selective observation" in transcripts. Liddicoat (2007) calls the act of transcription "an open-ended process in which the transcript changes as the researcher's insights into the talk are refined from ongoing analysis" ( 13). This implies that a transcript is not a static, objective tool, but rather an initially subjective one which

[109] Other technological issues included difficulties in synchronizing digital audio from two different sources, and also manufacturers' use of proprietary audio/video formats which makes manipulation difficult.

can change over time, depending on the researcher's direction and criteria. Liddicoat (ibid) goes on to discuss the issue of balance between detail and understandability. He states that when the subject matter is limited to *what* is said then transcripts are, for the most part, easy to read; when *how* something is said is transcribed, suddenly things become more complicated (ibid). For a multimodal approach using video data, the presence of gaze and gesture and facial expressions make written transcripts difficult to create and manage, and introduce a greater possibility for error. One example of a transcript that includes both recipient and speaker gaze is in Szatrowski (2002). Szatrowski's gloss makes use of numerous textual symbols to render the participants' action into text, transforming the original video data – something understandable to any speaker of the language, at least on a basic level – to something very complex, requiring a great deal of effort and background knowledge. At this point the transcript fails to be a reductionist tool and adds immeasurable complexity to the task at hand, while becoming increasingly questionable in its accuracy.

The question of what to transcribe is an issue that has not really been dealt with satisfactorily. Gail Jefferson (2004), whose transcripts arguably revolutionized the transcription process and the methods of language analysis in general, does not offer clear advice. Regarding what is found in her own transcripts, she writes: "… it's there because it's there[110], plus I think it's interesting" (15).  A few pages later, however, she mentions how inaccuracy (in terms of not including timing, prosodic features, overlap, etc.) in a transcript can change a researcher's interpretation of the data (pp. 16-17)[111].

[110] I assume this to mean, "it's there because it's there *in the conversation*."

[111] For this study I decided to code the data based on time and elements of interaction rather than create a transcript. While there are many kinks to work out I believe that using transcripts from the start may have affected my ability to objectively see everything that was going on in the data.

Can any reductionist transcript be completely accurate? Is a transcript that includes every element of the data necessary (not to mention useful or approachable) when the original recording is available?

## 5.6 Further implications

It is clear that much remains to be done on the topic of Japanese backchannels/*aizuchi*. I focused mainly on one kind of upward head movement, the 'thinking face,' and briefly discussed the 'head twist,' but there are many other bodily actions in this corpus alone that have yet to be described, including brief facial expressions such as frowning, pursed lips, and smiling; head shakes, laughter in the absence of substantial talk, and even a case of perfectly synchronized head nods performed by a speaker and a recipient in the complete absence of talk. I also believe that more investigation into 'typical' recipient responses, such as short verbal responses and vertical head nods, using a multimodal approach is necessary, since no study I know has looked at them within the full context of interaction.

Ford, Thompson, and Drake (2009) found that the wide variety of nonverbal devices utilized by both speakers and recipients sometimes made it difficult to fit their observations on English conversations into the established turn-based system. I found the same holds for Japanese. I also found that the amount of 'facially motivated collaboration' (smiling together, copying the speaker's expressions) that goes on obviously has a great effect on how the talk plays out, but accounting for it remains difficult, since it sometimes seems to have no direct relationship to the talk at hand. But despite the number of unanswered questions that remain, the results of this study point demonstrate that gaze has a large role to play in Japanese interaction, much as it does in English (Goodwin 1980), and that the effective analysis of video data is necessary to further our knowledge of previously overlooked behaviours that are so prevalent in and relevant to talk.

# 6 Works Cited

Aoki, H. (2008). *Hearership as Interactive Practice: A Multi-modal Analysis of the Response Token Nn and Head Nods in Japanese Casual Conversation.* PhD Dissertation, University of California, Los Angeles.

Aoki, H. (To appear). Some Functions of Speaker Head Nods in Japanese Casual Conversation. In C. Goodwin, H. Kim, & J. Streeck (Eds.), *Multimodality and Human Activity: Research on Human Behavior, Action, and Communication (working title).* Cambridge University Press.

Chafe, W. (1994). *Discourse, Consciousness, and Time.* Chicago: The University of Chicago Press.

Chafe, W. (Director). (1975). *The Pear Film* [Motion Picture].

Chafe, W. (1980). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production.* Norwood, NJ: Ablex.

Delancey, S. (2001). *On Functionalism.* Retrieved January 20, 2010, from DeLancey's Linguistics Home Page: http://www.uoregon.edu/~delancey/sb/fs.html

Erbaugh, M. S. (2001). Retrieved March 15, 2006, from The Chinese Pear Stories - Narratives Across Seven Chinese Dialects: http://pearstories.org/

Ford, C. E., Fox, B. A., & Thompson, S. A. Constituency and the Grammar of Turn Increments. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The Language of Turn and Sequence* (pp. 15-38). New York: Oxford University Press.

Ford, C. E., Thompson, S. A., & Drake, V. (2009, September 5-6). Turn Continuation and Bodily-visual Action. *3rd Turn Continuation Workshop in Cross-Linguistic Perspective* . University of Alberta.

Furo, H. (2001). *Turn-taking in English and Japanese: Projectability in Grammar, Intonation, and Semantics.* (L. Horn, Ed.) New York: Routledge.

Gardner, R. (2001). *When Listeners Talk: Response Tokens and Speaker Stance.* John Benjamins.

Givon, T., & Dickinson, C. (1997). Memory and Conversation. In T. Givon (Ed.), *Conversation: Cognitive, Communicative, and Social Perspectives* (pp. 91-92). Philadelphia, PA: John Benjamins.

Goodwin, C. (1981). *Conversational Organization.* Toronto: Academic Press.

Goodwin, C. (1980). Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry , 50* (3-4), 272-302.

Goodwin, C. (1979). The Interactive Construction of a Sentence in Natural Conversation. In G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology* (pp. 97-121). New York: Irvington Publishers.

Hatasa, Y. A., Hatasa, K., & Makino, S. (2011). *Nakama 1.* Boston: Heinle Cengage Learning.

Hayashi, M. (2003). *Joint Utterance Construction in Japanese Conversation.* Philadelphia: John Benjamins.

Iwasaki, S. (2009). Initiating Interactive Turn Spaces in Japanese Conversation: Local Projection and Collaborative Action. *Discourse Processes , 46:2*, 226-246.

Iwasaki, S. (1997). The Northridge Earthquake Conversations: The floor structure and the 'loop' sequence in Japanese Conversation. *Journal of Pragmatics , 28*, 662-693.

Iwasaki, S. (1993). The Structure of the Intonation Unit in Japanese. *Japanese Korean Linguistics , 3*, pp. 39-53.

Jefferson, G. (2004). Glossary of Transcript Symbols with an Introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies From the First Generation* (pp. 13-31). Philadelphia: John Benjamins.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters.* New York: Cambridge University Press.

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.

Lebra, T. S. (1976). *Japanese Patterns of Behavior.* Honolulu: University of Hawai'i Press.

Lee, Y. (2010). *The Ideology of Kokugo: Nationalizing Language in Modern Japan.* (M. H. Hubbard, Trans.) Honolulu: University of Hawai'i Press.

Linell, P. (2005). *The written language bias in linguistics: its nature, origins and transformations.* London: Routledge.

Maynard, S. K. (1990). Conversation Management in Contrast: Listener Response in Japanese and American English. *Journal of Pragmatics , 14*, 317-412.

Maynard, S. K. (1997). *Japanese Communication: Language and Thought in Context.* Honolulu: University of Hawaii Press.

Maynard, S. K. (1989). *Japanese Conversation: Self-contextualization through Structure and Interactional Management.* Norwood, NJ: Ablex Publishing.

McClave, E., Kim, H., Tamer, R., & Mileff, M. (2007). Head movements in the context of speech inn Arabic, Bulgarian, Korean, and African-American Vernacular English. *Gesture , 7* (3), 343-390.

Miller, L. (1983). *Aizuchi: Japanese Listening Behavior.* M.A. Thesis, Anthropology, UCLA.

Miller, R. A. (1982). *Japan's Modern Myth: The Language and Beyond.* New York: Weatherhill.

Nakane, I. (2007). *Silence in Intercultural Communication: Perceptions and Performance.* John Benjamins.

Ochs, E. (1979). Transcription as Theory. In E. Ochs, & B. B. Schieffelin (Eds.), *Developmental Pragmatics* (pp. 43-72). New York: Academic Press.

Sacks, H. (1992). *Lectures on Conversation.* (G. Jefferson, Ed.) Oxford: Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language , 50* (4), 696-735.

Schegloff, E. A. (1982). Discourse as an Interactional Achievement: Some Uses of 'Uh huh' and Other Things That Come Between Sentences. In D. Tannen (Ed.), *Analyzing Discourse: Text and Talk (Georgetown University Roundtable on Languages and Linguistics 1981)* (pp. 71-93). Washington, D.C.: Georgetown University Press.

Schegloff, E. A. (1990). On the Coherence of Sequences as a Source of "Coherence" in Talk-in-Interaction. In B. Dorval (Ed.). Norwood, New Jersey: Ablex Publishing Corporation.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In J. Lee, & G. Button (Eds.), *Talk and Social Organization* (pp. 70-85). Clevedon: Multilingual Matters.

Streeck, J. (1993). Gesture as Communication I: Its Coordination with Gaze and Speech. *Communication Monographs , 60* (4), 275-299.

Sugito, S. (1989). 言葉のあいづちと身ぶりのあいずち：談話行動における非言語的表現 "Kotoba no aizuchi to miburi no aizuchi: danwa koodoo ni okeru higengoteki hyoogen". *Nihongo Kyouiku , 67*, 48-59.

Szatrowski, P. (2002). Gaze, Head Nodding, and Aizuti in Information Presenting Acvtivities. In P. M. Clancy (Ed.), *Japanese/Korean Linguistics 11* (pp. 119-132). Stanford: CSLI.

Tanaka, H. (2000). The Particle ne as a Turn-management Device in Japanese Conversation. *Journal of Pragmatics , 32* (8), 1135-1176.

Tanaka, H. (1999). *Turn-taking in English and Japanese: A Study in Grammar and Interaction.* Philadelphia: John Benjamins.

Tanaka, L. (2004). *Gender, Language and Culture: A Study of Japanese Television Interview Discourse.* Philadelphia, PA: John Benjamins.

Tannen, D. (1993). *Framing in Discourse.* Toronto: Oxford University Press.

The Prague Linguistic Circle. (1982). Theses Presented to the First Congress of Slavic Philologists in Prague, 1929. In P. Steiner (Ed.), *The Prague School: Selected Writings, 1929-1946* (pp. 3-31). Austin: University of Texas Press.

Ward, N., & Tsukahara, W. (2000). Prosodic Features which Cue Back-channel Responses in English and Japanese. *Journal of Pragmatics , 23*, 1177-1201.

Yamada, H. (1992). *American and Japanese Business Discourse: A Comparison of Interactional Styles.* Norwood, N.J.: Ablex.

Yngve, V. H. (1970). On Getting a Word in Edgewise. *Papers from the Sixth Regional Meeting: Chicago Linguistic Society* (pp. 567-578). Chicago: Chicago Linguistic Society.

# 7 Appendices

## 7.1 Appendix 1 – Transcription symbols

| Symbol | Meaning | Symbol | Meaning |
|---|---|---|---|
| . | Falling intonation/grammatical ending. | , | Short pause |
| (#.#) | Pause (seconds) | <action> | Description of action beginning at point < and ending with > |
| <Gaze up/down> | Change in direction of gaze | -------------- | Continuation of action |
| [ | Point of overlapped speech | ? | Rising intonation |
| \ | Falling intonation (only marked where relevant to analysis) | = | Latching |

## 7.2 Appendix 2 – Equipment

I used a Toshiba Gigashot GSCR60 digital hard disk video camera for all recordings. Sound was recorded with a Marantz Professional PMD660 audio recorder in stereo WAV format at 44.1kHz/16 using the internal microphone, or using an Oade Brothers Audio modified Marantz PMD660 with a mono Oktava MK-012 microphone.

I used a variety of free software packages for the data preparation and analysis. Audacity was used for all audio editing & splicing. AoA Audio Extractor was used to extract audio from the video files. VirtualDub was used for video editing and encoding. GOM Player was used for viewing the video files and for capturing video stills. I used IrfanView for basic editing and preparation of the video stills.