FORMS AND PROSODIC CHARACTERISTICS OF BACKCHANNELS IN TOKYO AND OSAKA JAPANESE

Yasuhiro Katagiri *, Miyoko Sugito †, and Yasuko Nagano-Madsen ‡
*ATR Media Integration & Communications Research Laboratories, Japan
†Institute for Speech Communication Research, Japan
‡University of Gothenburg, Sweden

ABSTRACT
Linguistic forms, prosodic characteristics and conversational functions of Japanese backchannels ('aiduti') were investigated. Based on an empirical analysis of recorded conversations in the Tokyo and Osaka dialects of Japanese, a bipartite classification scheme is proposed for backchannel items in terms of their types of linguistic forms. With the scheme, it was found that the first category of backchannels constitutes the major use of backchannels across the two dialects and social groups, whereas the dialectal and sociolinguistic variations in backchannel forms are confined to the second category. An experiment was then conducted, with the help of speech processing technology, to assess conversational functions of forms and prosodic characteristics of backchannel utterances. It was found that the use of backchannels and their appropriate temporal placements are significant in making a lively and smooth conversation.

1 INTRODUCTION
Backchannels are ubiquitous in human conversations. They serve to establish common grounds among conversants by sending back acknowledgments to the speakers [8]. They help coordinate turn-taking behaviors by passing up the opportunities of taking turns [6]. They also serve to create a sense of participation in a joint conversational event [4, 7], which is essential in sustaining a successful conversational interaction.

The present study reports on an empirical analysis and an experimental analysis of linguistic/prosodic characteristics of Japanese backchannels ('aiduti'), and their relevance to conversational functions. After a brief introduction on previous studies and conceptions on Japanese backchannels in section 2, we present an empirical analysis of backchannels based on recorded conversations in the Tokyo and Osaka dialects of Japanese in section 3. We propose a classification scheme for backchannels according to their linguistic forms, and discuss dialectal and socio-linguistic variations. In section 4, we present an experimental analysis of conversational functions of Japanese backchannels. Linguistic and prosodic characteristics of backchannels are artificially modified, with the help of speech processing technology, and their conversational effects are investigated.

2 JAPANESE BACKCHANNELS
It has been pointed out for some time that Japanese is a language in which so-called backchannels are used more frequently than in many other languages (for more details, see Horiguchi's overview on backchannels in Japanese [1]).

Maynard showed in her study that verbal backchannels in Japanese conversations were 2.9 times more frequent than in similar English conversations by Americans [3]. The use of backchannels is also taught at an early stage in Japanese language education for foreigners, as it is considered to be "very Japanese". These facts indicate that backchannels are one of the most eminent features of communications in Japanese.

All of the studies on Japanese backchannels so far, however, have dealt with a relatively limited scope, and such questions related to whether or not there exists a significant difference among the dialects of Japanese, how backchannels can be classified, what prosodic features they may exhibit, and how linguistic and prosodic characteristics of backchannels contribute to the conversational processes are still widely open questions.

Backchannels are defined as a listener's response to a signal that he/she is listening to and is willing to listen to more. In other words, backchannel utterances are thought to be rather contentless without any substantial or new information. Such utterances as [hai] (=yes) are each disregarded as a backchannel if it had been used as an answer to a speaker's question or command.

Backchannels can also include non-verbal behaviors such as nodding, laughter and gestures. In the present study, only audible laughters which were marked in transcription were subjected to analysis.

3 AN EMPIRICAL ANALYSIS OF JAPANESE BACKCHANNELS
3.1 Introduction
First, an empirical analysis was conducted on linguistic forms and prosodic characteristics of Japanese backchannels by inspecting actual Japanese conversational data. The analysis was focused on variations in forms and usage frequencies of backchannels in different dialects and in different social groups.

3.2 Material
The data used for the present analysis were recordings and broad transcriptions of four dialogues by native speakers of the Tokyo and Osaka dialects (Tokyo uptown and downtown; Osaka uptown and downtown) of Japanese, which were recorded by researchers at the National Language Research Institute. Each dialogue consisted of four middle-aged or elderly speakers, and sociolinguistic factors such as gender and intimacy between speakers had been taken into...
consideration in organizing the group. In addition, a researcher joined the group as a chairperson to give rough guidance on the topics of the dialogues. The first 45-minute portion was taken from each of the four dialogues and subjected to the analysis, making the total analysis time approximately 3 hours.

3.3 Method

Backchannel utterances were picked out according to the above definition through careful listening of the recordings of the dialogues. Backchannel items were marked in the transcription, and then collected by their linguistic forms. Fundamental frequency (F0) analysis was also performed on selected backchannel utterances.

3.4 Results

3.4.1 Classification. Altogether 2,870 backchannel items were extracted. The number of backchannels used for each group was 574 (Tokyo uptown), 580 (Tokyo downtown), 319 (Osaka uptown) or 664 (Osaka downtown).

These backchannels were first classified according to their basic linguistic forms as the main criteria. This classification resulted in 11 classes: {/Ni:/, /aso:desuka/, /a:/, /e:/, /hai/, "laughter," "echo (repetition)," /haheho/, /so:/, /ne/ and "others." Most of these appeared either long or short as in [so] and [so:]. Further analysis revealed that a large number of backchannels were formed through a process called reduplication as in [c][c][c][c][c][c] or in [hai][haihai][haihaihai]. Based on a reduplication criterion, the 11 classes were further divided into two main categories: (1) one that can form its reduplicated counterpart and (2) the other, which usually doesn’t. Category (1) consisted of /Ni:/, /e:/, /a:/, /so:/, /hai/, /haheho/ and "laughter," while group (2) consisted of /ne/, /aso:desuka/ (echo (repetition)) and "others."

3.4.2 Usage patterns. The usage frequency analysis revealed that the large majority of backchannels in Japanese belong to category (1). 2,167 backchannel instances out of 2,870 (75.5%) were of the category (1) in our data. The number of different types of backchannels used by each speaker ranged from 8 to 11. This indicated that Japanese speakers share a common repertoire of backchannel expression types of a size of approximately 10, and possibly assign different functions to different types of backchannels. An interesting finding was that forms that undergo dialectal and sociolinguistic variations were concentrated in the items in category (2), some of which looked like more substantial utterances. /Ne/ in the Tokyo dialect becomes /na/ in the Osaka dialect, and /aso:desuka/ in male speech becomes /aso:desuka/, a politier form, in female speech. It follows then that typical backchannels in Japanese are those with reduplicated variants for which very little variations occur due to dialectal, gender and intimacy differences. A distribution involving the use of backchannels in different groups is summarized in Table 1.

The reduplication of backchannels primarily indicates emphasis in the listener’s responses. The exact magnitude of reduplication as well as frequency of occurrence for reduplicated variants differ across backchannel classes as well as the dialogue contexts in which they occur. Out of the 1,669 tokens of the category (1) backchannels (excluding /haheho/ and “laughter”) in our data, 12% had reduplicated forms. A typical reduplicated form had double reduplications as in [c][c] and [haihai], while those with more than three reduplications were rare. The exception to this was the /so/ class for which it was more common to use longer reduplicated forms. For a more detailed analysis of backchannels in Tokyo and Osaka Japanese, readers are referred to Nagano Madsen and Sugito [5].

3.4.3 Prosodic characteristics. Since it was not easy to extract clear F0 contours from dialogues in which five speakers participated, only some of the F0 tokens of a male speaker from the Tokyo uptown group were analyzed for more detailed measurements. A typical intonation pattern used for the backchannels was gradual falling (III). It was observed that F0 contours representing different backchannels were extremely similar regardless of their sound structures. Even longer utterances with reduplications exhibited similar patterns. The only exceptions were variants of /N/ for which a somewhat lower pitch range was occasionally used.

The values of F0 maxima were also measured for the utterances of the Tokyo uptown male speaker. It was observed that this speaker used three major pitch ranges in which the F0 maxima occurred: 180–200 Hz (high), 100–120 Hz (mid) and around 90 Hz (low). Among the three pitch ranges, the

<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Tokyo</th>
<th>Osaka</th>
<th>Chairperson</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>/Ni:/</td>
<td>145</td>
<td>201</td>
<td>37</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>48</td>
<td>59</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>/a:/</td>
<td>45</td>
<td>48</td>
<td>54</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>/so:/</td>
<td>13</td>
<td>18</td>
<td>11</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>/hai/</td>
<td>63</td>
<td>37</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>/haheho/</td>
<td>3</td>
<td>7</td>
<td>43</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>laughter</td>
<td>126</td>
<td>63</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>(2)</td>
<td>/ne/</td>
<td>9</td>
<td>19</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>/aso:desuka/</td>
<td>82</td>
<td>74</td>
<td>19</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>echo (repetition)</td>
<td>37</td>
<td>44</td>
<td>22</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>others</td>
<td>2</td>
<td>10</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>574</td>
<td>580</td>
<td>319</td>
<td>664</td>
</tr>
</tbody>
</table>

Table 1: Use of backchannels in Japanese conversation.
most frequently used one was the mid pitch range. The low pitch range was used mostly for variants of /N/, while the high pitch range was used mostly for longer utterances with many reduplications as in /so:so:so:so:/ This indicated that the high pitch range and the reduplications of backchannels worked together to indicate stronger listener responses, thereby contributing to a livelier conversation flow.

4 CONVERSATIONAL EFFECTS OF JAPANESE BACKCHANNELS

4.1 Introduction
In order to further elucidate the functions of backchannels serve in conversations, an experiment was conducted, comparing the effects of the linguistic forms and prosodic characteristics of the category (1) backchannels in conversations. Variations of the conversational materials were electronically created out of excerpts of natural conversation data. Speech analysis, modification and synthesis program STRAIGHT [2] was employed to manipulate speech characteristics of the backchannels and to obtain high-quality synthesized speech. The manipulated conversational materials were then subjected to a listening test to find out what impressions they gave to evaluate the comparative effects of the manipulations on conversations.

4.2 Material
The original dialogue data for the experiment were taken from a spoken dialogue corpus collected at ATR Media Integration & Communications Research Labs. Four dialogues were chosen to include all of the combinations of the dialectal variations between the Tokyo and Osaka dialects of Japanese, and the conversation style variations between task-oriented dialogues and free conversations. Portions of approximately 40 seconds containing a good number of backchannel utterances were taken from each of the dialogues, and were subjected to subsequent backchannel manipulation operations.

4.3 Method
The backchannel utterances in each of the four dialogue data were manipulated by the following procedure:
1. First, segment speech at each stretch of silence into a sequence of inter-pausal units. Perform the segmentation separately for each speaker.
2. Mark the backchannel items according to the definition of backchannels by referring to the transcriptions.

3. Analyze the entire speech data with STRAIGHT speech processing package.
4. Subject the backchannel portions of the speech to modification according to the five conditions shown below.
5. Synthesize target speech materials were with the modified parameters.
The speech manipulation parameters were chosen so as not to produce obviously artificial and non-natural speech. The conditions of speech modifications were:
Plain: No manipulation was performed on the speech.
Delay: Backchannel utterances were delayed by 600 msec relative to the preceding utterances, and, conversely, the following utterances by the partner were shifted earlier in time by 200 msec.
Pitch: Pitch patterns of backchannel utterances were modified so as to make slightly rising intonation patterns, by linearly incrementing $F_0$ toward the end of each backchannel utterance until it becomes 50% higher than the original value.
Form: Since merely substituting an occurrence of a backchannel utterance with a different backchannel occurrence, even by the same speaker, may not preserve its prosodic characteristics, and hence can produce effects not originated by linguistic forms, we chose to substitute each occurrence of a backchannel with its duplicated form, which was constructed by repeating the original backchannel utterance.
Removed: All backchannel utterances were removed and substituted with silence
All of the five manipulations were performed on the four dialogues, resulting in 20 conversation materials.

In the experiment, all 20 conversation materials were presented in random order to a group of subjects, who were then asked to score their impression on each conversation material in terms of the following five measures: (1) liveliness of conversation, (2) smoothness of conversation flow, (3) conversant involvement, (4) intimacy between conversants and (5) subject interest. Each measure consisted of a 7-point scale, from 1 (low) to 7 (high). A total of 14 subjects, 6 females and 8 males, participated in the experiment.

4.4 Results
4.4.1 Dialectal and style differences. An analysis of variance was performed on the scores obtained from the experiment. The factors for the analysis were dialect, conversation style and speech manipulation. The effects of dialectal and conversation style differences were examined first.
Significant main effects were observed for the dialect factor in the liveliness measure and the involvement measure. Conversations in the Osaka dialect were scored significantly higher ($p < .01$) than those in the Tokyo dialect in these two measures, confirming the popular stereotype that Osaka people are more talkative than Tokyo people (Figure 1). A significant main effect was also found for the conversation style factor in the subject interest measure. Free conversations were judged to be more interesting ($p < .01$) than task-oriented dialogues by the subjects (Figure 2). These results agreed with our intuitive estimates on dialectal and conversation style variations, and hence confirmed, albeit indirectly, that our speech manipulation operations were kept within reasonable bounds so as to preserve the natural quality of the original conversations.

4.4.2 Speech characteristics. Significant main effects were observed for the speech manipulation factor in all of the five measures. The effects of the speech characteristics manipulations on each of the five measures are shown in Figure 3. The figure shows that, for all five measures, the Removed and Delay conditions scored lower than the other three conditions. A pairwise $t$-test confirmed that the Removed and Delay conditions were scored significantly lower ($p < .05$) than the Plain condition for all of the five measures. Consequently, we can say that conversations, when lacking backchannels, are dull, awkward, gloomy, distant and of less interest. Even delayed backchannel responses create similar impressions.

In contrast to this, the Pitch and Form conditions did not show significant differences from the Plain condition for any of the five measures. Neither did the reduplication of backchannels nor a rising intonation on backchannel utterances give strong effects on the overall conversation impressions. It should be noted, however, that the reduplication of and rising intonation of backchannels produced slightly higher scores in the liveliness and involvement measures. All of this suggested that, rather than exercising a universal function which sets the overall impressions of conversations, both linguistic form and intonation of backchannels contribute to the backchannels on an individual basis, and determine the conversational functions of the backchannels depending on the contexts in which they occur.

No interactions were observed between either the dialect factor or the conversation style factor and the speech manipulation factor. This indicated that conversational functions of backchannels are universal across dialects and also across conversation styles.

5 CONCLUSIONS

An investigation was conducted on the usage and the functions of Japanese backchannels ('aidutī') in conversations. A classification of backchannel items was proposed based on the formability of reduplications. An examination of actual Japanese conversation data revealed that backchannel items able to be reduplicated constitute the major category of backchannels, whereas dialectal and socio-linguistic variations are confined to backchannel items not able to be reduplicated. A further experimental study involving the artificial manipulation of speech characteristics of backchannel utterances in conversations demonstrated that the use of backchannels and their appropriate temporal placements are essential in making conversations lively and smooth. It was also suggested that contributions of linguistic forms and intonation patterns to backchannel functions are dependent upon the contexts of use. Further investigations are needed to clarify these context dependent functions, and joint effects of forms and prosodic characteristics of backchannels.

ACKNOWLEDGMENTS

Part of the research reported in the paper was conducted while the third author was at the Institute of Speech Communication Research as a fellow of the Japan Foundation. The authors thank Yoshimori Sakane, Miwako Kurihara and Yugo Takeuchi for their suggestions and support in conducting the experiment.

REFERENCES