Japanese children’s acquisition of prosodic politeness expressions

Takaaki Shochi 1, Donna Erickson 2, Kaoru Sekiyama 1, Albert Rilliard 3 & Véronique Aubergé 4

1 Division of Cognitive Psychology, Kumamoto University, Japan
2 Showa Music University, Kawasaki City, Japan
3 LIMSI-CNRS, Orsay, France
4 GIPSA Lab, Grenoble, France

shochi38@gmail.com, ericksondonna2000@gmail.com, sekiyama@kumamoto-u.ac.jp, albert.rilliard@limsi.fr, veronique.auberge@gipsalab.inpg.fr

Abstract

This paper presents a perception experiment to measure the ability of Japanese children in fourth and fifth grade elementary school to recognize culturally encoded expressions of politeness and impoliteness in their native language. Audiovisual stimuli were presented to listeners, who rate the politeness degree and a possible situation where such an expression could be used. Analysis of results focuses on the differences and the similarities between adult listeners and children, for each attitude and modality. Facial information seems to be retrieved earlier than audio ones, and expressions of different degrees of Japanese politeness, including expressions of kyoshuku, are still not understood around 10 years of age.

Index Terms: prosodic social affects, language acquisition, Japanese.

1. Introduction

In face-to-face interaction, social affects (or attitudes) are expressed within the multimodality of speech [1]. Since attitudes are highly linked to language and culture, various attitudinal expressions are supposed to be learned in the surrounding environment [2]. In particular, verbal politeness expressions are strongly influenced by the culture of the speaker [3].

This study examines how well Japanese children from 9 to 11 years old can identify prosodically expressed politeness attitudes in the Japanese language: kyoshuku politeness, sincerity-politeness, simple-politeness, declaration (politely neutral) and arrogance, from an utterance which contains an affectively neutral meaning. Among these social affects, sincerity-politeness and kyoshuku are conventionalized in Japanese society as a type of politeness expression used to convey the speaker’s sincerity (for sincerity-politeness) or ashamedness / embarrassment (for kyoshuku) to the hearer in order to avoid a conflict situation or to get a favor from a hierarchically superior person [4].

Previous studies show that these affects are signaled by various acoustic cues, such as $F_0$ characteristics (intonation contour, $F_0$ height and amount of change in $F_0$), loudness, duration and voice quality [5]. They also indicate that the gender and age of the listener as well as the language background influences perception of affective expressions ([6], [7], [8], [9]). Moreover, our previous research showed the important indices of facial muscles’ movements independently to the acoustic cues for social politeness expressions [10]. However, only a few studies have examined the perceptual behavior of children for expressions of social affects in multimodal conditions. We aim here at measuring the perceptual relevance for Japanese children of audio-visual prosodic cues to the conventional expressions of Japanese politeness. Their performances are compared to those of adults on the same task.

Section 2 presents the experimental setup, the corpus used and the listeners involved in the test. Section 3 describes and analyses the results, and section 4 gives a summary of the main findings and their implications.

2. Experimental setup

Our previous work ([3], [10]) described the contribution of audio and visual information for 12 Japanese social affects as perceived by Japanese adult listeners. Perception results show that native listeners are able to recognize the 12 Japanese attitudes, while American English and French listeners show important confusions between some expressions. More specifically, the kyoshuku expression is not recognized as a politeness expression by foreign listeners, but as irritation or arrogance while the other politeness expressions receive better recognition scores [11]. The motivation of this work was thus to investigate this dimension of politeness expressions in Japanese, in order to measure if and when Japanese children start to recognize cultural expressions as adults do.

2.1. Corpus

Five Japanese social affects were examined, ranging from polite to impolite expressions: kyoshuku (KYO), sincerity-politeness (SIN), simple-politeness (PO), declaration (DC, politely neutral) and arrogance (AR). Each audio-visual stimulus was recorded by a trained Japanese language teacher using the same affectively neutral sentence to express the prosodic attitudes [11]. They were presented to listeners either with the audio-only, video-only or audio-video modalities. The test consisted of 15 different stimuli (5 attitudes * 3 modalities) presented 3 times to listeners, in a random order.

2.2. Listeners

3 groups of listeners took the experiment according to age/grade in school. The first group consisted of 29 adults (mean age = 23, 18 females, 11 males); the second of 28 children in the 4th grade (G4) in the Japanese school system (mean age = 9.9 years, 13 females, 15 males); the third of 50 children in the 5th grade (G5) (mean age = 10.9 years, 31 females, 19 males). The total number of subjects for this experiment was 107. Adults took the experiment in a quiet room individually using a computer and headset. Children took the experiment in a school classroom with a computer and headset for each child. The task was described and after explanations and questions, they put their headsets on and took the experiment. A few listeners did not complete the experiment (they are not included in the above description).

Copyright © 2009 ISCA

1743
6 – 10 September, Brighton UK
No listener reported any perception trouble. The difference in size of each group is due to the fact that the children took the test as a class and the class size varied.

2.3. Paradigm
Stimuli were presented to the subjects in three experimental conditions, according to the modality of presentation: audio-only (A), video-only (V), audio-video (AV). The first two conditions are audio-only and video-only, and listeners perceive audio-video stimuli in the third one. In order to balance the effect of the presentation order, half the subjects began with the video-only stimuli, the other half with the audio-only ones. Each stimulus could be listened to only once.

After the presentation, subjects answered the question: "Who may this person talk to in this way?" The possible answers are: "Teacher", "Higher grade student/person", "Lower grade student/person", "Classmate" and "Unknown". This question was intended to check the kind of interpersonal relationship associated with the different attitudes by the subject within a hierarchical social framework that children can easily understand. They were then asked to judge the degree of politeness of the stimulus on a free scale ranging from impolite to polite (encoded on a 0-10 scale); the middle of the scale corresponded to a neutral expression. The position was set to the middle of the scale for each presentation, and subjects had to change it accordingly.

Each subject took the three experimental conditions (corresponding to the three modalities of presentation) during the same sitting. All tests were run on a computer interface, and listeners wore quality headsets.

3. Results

3.1. Statistical processing
Results given by subjects to the two questions were analyzed separately. The first question (the category of people who may be addressed in such a way by the speaker) was analyzed using correspondence analysis (CA). On the basis of contingency tables counting the number of answers in each category of answer – "Teacher" (Prof.), "Higher grade student/person" (HiGr), "Lower grade student/person" (LoGr), "Classmate" and "unknown" – for each presented attitude, the proximity of attitudes and of social status was measured. Different analyses were carried out for each of the three levels of subjects (adults, G4 and G5), mixing answers of the three modalities.

The answers to the second question – the degree of politeness – were analyzed using a GLM repeated measure procedure of SPSS. There were three within-subject factors: the presented attitudes (A - 5 levels, fixed factor), the modality of stimuli (M - 3 levels, fixed factor) and the three repetition of each stimulus (R - 3 levels, fixed factor). Two between-subjects factors were used: age of subjects/level in school (4th grade, 5th grade, or adult) (L - 3 levels, fixed) and the order of presentation of modalities during the test (O - 2 levels, fixed). Following [12], the Huynh-Feldt correction was applied to all results of the repeated measures ANOVA, in order to avoid possible deviations from the sphericity condition.

3.2. Description of results

3.2.1. Question 1
Results for question 1 for each group of subjects are summarized in figure 1; it shows the plot of the two first dimensions of the CA run for each group. For each group, the first two dimensions represent more than 95% of the total variance. The main similarity is the systematic association by all subjects of the answer "Lower grade student/person" to the expression of arrogance; and the polite expressions as well as the neutral one are on the other side of the first dimension. Amongst these polite expressions, adults clearly associate kyoshuku to "Higher grade student/person", simple politeness and sincerity-politeness to "Teacher", while the neutral declaration is associated to "Unknown". No systematic associations are made with "Classmate".

Both G4 and G5 children tend to associate politeness with "Teacher"; however, G4 associate kyoshuku with "Classmate",
and G5 with “Higher Grade student”, similar with what adults do. These results suggest that G4 students still do not have a complete grasp of kyoshuku.

Table 1. Results of the repeated measures ANOVA on the degree of politeness. Only significant factors at the 1% level are reported.

<table>
<thead>
<tr>
<th>Source</th>
<th>F</th>
<th>df</th>
<th>df error</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude</td>
<td>199.64</td>
<td>3.2</td>
<td>322.0</td>
<td>0.000</td>
</tr>
<tr>
<td>A*L</td>
<td>4.52</td>
<td>6.4</td>
<td>322.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Modality</td>
<td>8.07</td>
<td>2</td>
<td>202</td>
<td>0.000</td>
</tr>
<tr>
<td>A*M</td>
<td>8.19</td>
<td>8</td>
<td>804.4</td>
<td>0.000</td>
</tr>
<tr>
<td>A<em>R</em>O</td>
<td>2.90</td>
<td>8</td>
<td>808</td>
<td>0.003</td>
</tr>
</tbody>
</table>

3.2.2. Question 2

The ANOVA’s output is presented in table 1. The attitudes had a main effect on listeners’ answers, as well as the modality of presentation and the interaction between both factors. Moreover, the interaction between attitudes and the level of subjects (children at G4 or G5 or adults) was also significant. The interaction between attitude, repetition and the order of modalities presentation was significant, but seems difficult to interpret. Figure 2 presents the mean answer given by each category of subject to the attitudes in each modality.

Effect of attitude: Results clearly show the main effect of attitude on the answer. Arrogance was perceived as the most impolite expression (mean=2.7), while all the others received scores over 5. Simple-politeness was rated as the most polite expression (6.7), with sincerity-politeness (6.1). Declaration received a mean score of 5.8, while kyoshuku got a 5.

Interaction attitude*level is represented in Figure 3. Interestingly, the politeness degree of kyoshuku systematically rose with age, from 4.7 at G4, 5 at G5 to 5.3 for adults, while it decreased for declaration (G4=6.3; G5=5.9; AD=5.3). Arrogance was also better recognized by older subjects as an impolite expression (G4=3.3; G5=2.4; AD=2.6), while the two expressions of politeness didn’t change much with age.

Effect of modality: Visual information was linked with a decrease of politeness degree (A=5.4; V=5.2; AV=5.2), the effect of which was mainly linked with the importance of visual information for arrogance (see below).

Interaction attitude*modality: The audio-visual presentation allowed listeners to give the most accurate evaluation of each attitude. However, results for all modalities were similar across attitudes with the exception of arrogance, for which the visual information helped listeners to perceive the impolite expression (A=3.4; V=2.5; AV=2.1).

4. Discussion & Conclusions

This paper investigates the perceptual behavior of adults vs. children for five prosodic expressions of Japanese politeness. In previous research, prosodic parameters (F0, moraic duration and intensity) and facial movements (labeled with the FACS Manual [14]) of each social politeness expression were analyzed [10]: Arrogance is mainly linked with a low F0, lengthening, a high intensity and an activation of the inner brow raiser. Declaration is characterized by a low and decreasing F0 a final lengthening, a weak and decreasing intensity and (for this performance) a movement of the head up (and no other visual indices). Kyoshuku is performed with a flat F0, almost isochronic rhythm, low intensity, and activation of the brow lowerer, the eyes closed and the head down. Politeness is performed with a high pitch, a slow rhythm, a weak intensity, the head forward and a head nod. Finally, Sincerity-politeness is performed, as Politeness, with a high pitch, a slow rhythm, a weak intensity, the head forward and a head nod. Finally, Sincerity-politeness is performed, as Politeness, with a high pitch, no particular intensity pattern and the same facial expression, but they differ on rhythm (Politeness is slower than Sincerity-politeness).

The major finding of current perception results is that all listeners, regardless of age, can distinguish between impolite and polite due probably to the difference of intensity. However,
a complete adult-like understanding of Japanese social politeness is not yet achieved by G4 and G5 children. Instead, we see a progression of acquisition of levels of politeness. For G4 listeners, everything is polite except for kyoshuku. For G5 listeners, we see the beginning of awareness of differences among the various levels of Japanese politeness, even including kyoshuku. The identification of kyoshuku by G5 children is not as high as for adults, but the association of kyoshuku is now with HiGr student/person, as it is with adults—no longer with classmates.

An explanation why adults’ associate kyoshuku with HiGr person, but not with Prof. may be that this attitude is a type of social politeness that occurs only in intimate (or sometimes conflicted) relationships, depending on the speaker’s assessment of the situation. The kyoshuku expression does not really stand on a scale of politeness, understood as a politeness of courtesy, and this may explain why even adults did not rate it as a very polite expression but almost as politely neutral. It rather involves a knowledge of psychological distance between speaker and hearer, which implies the necessity of understanding the social context in which this attitude occurs. Why G4 children associate kyoshuku with classmate is interesting. It may be that G4 children use a kyoshuku type of expression in order to get a favor, or ask forgiveness, and in this sense, it is similar to the adult usage. But children can only use this expression with a classmate, of course not with a HiGr person; they still lack the notion of hierarchy associated with adult-like notion of kyoshuku.

Looking at the arrogance results, for all levels, the audio-only receives significantly lower scores of impoliteness than audio-video stimuli (Post-hoc tukey test, p<.001). Especially, this is so for G4 children who rate audio-only arrogance over 4 on the politeness level. However, arrogance is systematically associated with “Lower grade student/person”, whatever the age of the subject. Therefore all listeners seem to understand perfectly this notion and the way it is performed in Japanese culture – at least through facial expressions.

Opposite observations are made for the expressions of sincerity-politeness and simple-politeness rated by children: politeness degree is higher when judged from audio (and audio-visual) cues (performed with a higher F0 and slower tempo for politeness, which is coherent with the “frequency code” proposed by [13]) than from facial ones. These findings are consistent with the results obtained by [11] on Japanese attitudes, with the expressions of simple-politeness and sincerity-politeness mainly expressed through audio cues, whereas arrogance relies more on visual cues.

The results also suggest that neither G4 nor G5 children understand the Japanese adult concept of the differences between sincere-politeness, simple-politeness and declaration. G4 children group sincere-politeness with declaration and G5 children seem to realize that sincere-politeness is different from declaration but have not yet associated it with the expression of politeness used with teacher, as adults do. For both G4 and G5 children, who do not discriminate accurately between declaration and forms of politeness, there is no significant difference due to modality of presentation. Conversely, adult listeners separate declarative from expressions of politeness more accurately with video-only and audio-video modalities than with audio-only modality (post-hoc Tukey test, p<.01).

This experiment on the perception of culturally encoded expressions of politeness and impoliteness in Japanese by children gives interesting results about both the age when they begin to be able to recognize some of these expressions, as well as about their preferred modalities to perform such a task. Politely neutral rating of kyoshuku by adult listeners also raises interesting questions about the conceptual dimension on which kyoshuku is expressed. It also reveals that the perceptual behavior of adult listeners for kyoshuku is affected by the psychological distance between the speaker and the receiver. It implies that the perception of this attitude requires the social context where it might occur.

To investigate further the development of cultural expressions’ recognition, other perceptions tests are currently being done with 13/14 year old children. We also plan to test younger children but since young children cannot read, problems of how to administer the test have to be solved. Interactive paradigms may be a solution using virtual expressive avatars for purposes of computer-child interaction.

5. Acknowledgements

We are deeply grateful to children of Kurokami and Izumoigaoka elementary schools as well as their professors and the directors of these schools, to allow us to present our experiments. We also thank the adult listeners who participate in this experiment, and especially, T. Ohmac. This work was supported by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010)19520371 to the second author, as well in part by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

6. References