# John Benjamins Publishing Company

Jb

This is a contribution from *Multimodal Im/politeness. Signed, spoken, written.* Edited by Andreas H. Jucker, Iris Hübscher and Lucien Brown. © 2023. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com). Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

# Perceptual changes between adults and children for multimodal im/politeness in Japanese

Takaaki Shochi<sup>1, 2</sup> Albert Rilliard<sup>3, 4</sup> & Donna Erickson<sup>5</sup> <sup>1</sup> CLLE UMR5263 CNRS, Pessac, France | <sup>2</sup> LaBRI UMR5800 CNRS, Talence, France | <sup>3</sup> Université Paris Saclay, CNRS, LISN | <sup>4</sup> Universidade Federal do Rio de Janeiro | <sup>5</sup> Haskins Laboratories

Japanese im/politeness attitudes are conveyed by audiovisual prosody in tandem with lexical markers. This chapter reports on two experiments about the acquisition by Japanese elementary school children of prosodic codes and social categories. The first experiment studied the perceived degree of politeness and its social use, and the second, the perceived similarity between the expressivities conveyed in pairs of expressions. An analysis of the audiovisual performances showed the types of changes in pitch and range in line with symbolic frequency and effort codes. The perceptual results showed that children learn to use and recognize im/polite expressions in a socially adequate fashion between 6 to 10 years old, thus showing an underlying growing cultural coherence gained with age.

**Keywords:** perceptual changes, multimodal expressivity, im/politeness, Japanese

# 1. Introduction

During human communication, multimodal cues are used for conveying and receiving social affective information and play an important role for establishing good interpersonal relationships (e.g., Swerts and Krahmer 2005; Swerts 2011; Rilliard et al. 2009, 2014; Erickson et al. 2020). Different from basic emotions (Izard 2007) which are probably shared among human beings as a phylogenetically developed competence (see Fernández-Dols and Russell 2017 for discussions), social affects, such as politeness, irritation, irony, contempt etc., are closely linked to language and culture. It follows, therefore, that these expressions can vary across languages and cultures, and as such their prototypical characteristics and socially-conformed use must be learned (Zinck and Newen 2008; Scherer 2013). Symbolic codes have been proposed in the literature that link the recurrent correlation between a physiologically derived constraint (the size of the vocal folds or the link between fundamental frequency and effort) and changes in the voice, and propose these relations are used to encode the reverse relationship: because larger animals tend to have lower voices than conspecifics, lowering one's voice sends a message of "larger size", and thus more power etc., under Ohala's Frequency code (1983, 1994). From this observation, he predicts the use of higher/lower pitch for a series of situations that include relations of power – and typically useful here, in the case of "politeness", a higher pitch would be predicted. Gussenhoven's Effort code (2004) links a greater effort of the speaker in a given speech act as a marker of its importance for the speaker's goals; greater effort leads to larger pitch span – and marking interest in an interaction is a potentially polite behavior, thus the prediction of larger pitch span following the Effort code.

Social affective expressions are closely linked to im/politeness strategies, and are thus determined by three things for the management of face: distance, relative power and absolute ranking (Brown and Levinson 1987; see also Culpeper, Bousfield, and Wichmann 2003; Culpeper 2010, 2016; or works in Culpeper, Haugh, and Kádár 2017, for impoliteness specificities). Distance is a social variable which describes a symmetric social dimension of proximity between the speaker and the hearer, in terms of similarity and frequency of interaction. Relative power indicates, still in Brown and Levinson's (1987) view, the amount of imposition the speaker can put on the hearer, as well as the self-evaluation of the social status difference between speaker and hearer, that may legitimate this potential to impose on / control the other (on distance and power, see also Spencer-Oatey 1996). Power is based on social hierarchy, which is what is seen in Japanese society. Absolute ranking refers to "a culturally and situationally defined ranking of impositions by the degree to which they are considered to interfere with an agent's wants of self-determination or of approval" (Brown and Levinson 1987: 77).

In Japanese society, the cultural concepts of politeness are highly codified by social status (rank), age and situation, and seem to fit well with Brown and Levinson's account of im/politeness strategies. The sociological variables relating to im/politeness strategies are strongly influenced by age and social hierarchy, culturally-determined factors which play a decisive role in establishing good relationships among conversation partners, that may account for the "discernment" aspect of social interaction in a high context culture like the Japanese one (Hill et al. 1986; Ide 2002). See Kiyama, Tamaoka, and Takiura (2012) for a detailed discussion of the applicability of Brown and Levinson's face model to Japanese culture, as a rebuttal to earlier studies, e.g. Ide (1989, 2006) and Matsumoto (1988), suggesting that Brown and Levinson might be problematic for Asian cultures.

Previous perceptual studies have examined perception of the vocal and visual characteristics of im/politeness strategies. Nadeu and Prieto (2011) have shown the importance of facial information to interpret the acoustic information linked to the expression of politeness in Catalan. Building on the "audio-visual prosody" paradigm (Swerts and Krahmer 2005), studies have shown that the visual cues have a role, together with audio and lexical ones, in the expression and identification process of many linguistic functions, notably in relation to the acquisition of social awareness in children (Swerts 2011), and capacities related to the management of linguistic im/politeness strategies. These studies include works studying the reception of im/politeness cues in listeners of various language backgrounds, for example French, English, Brazilian Portuguese, and Japanese in Rilliard et al. (2014). Also see Idemaru, Winter, and Brown (2019) for a comparison of Japanese and Korean polite speech, and Idemaru et al. (2020) for Korean deferential speech. Other studies showed variation in the reception of audiovisual prosodic cues to social affect cross-culturally (Mixdorff et al. 2017; Rilliard et al. 2017; Shochi, Aubergé, and Rilliard 2007).

Concerning studies of audio-visual prosody specifically with regard to first language acquisition, Swerts (2011) examined the relation between visual prosody and social awareness in children, reporting that as children become older, they become more socially aware, and this is expressed in their prosody. For Swerts, audio-visual prosody refers to both auditory cues as well as visual cues, such as body language, posture and facial expressions. With age, the child becomes "more knowledgeable about the conventions that exist within a specific community on how to associate specific expressions with specific pragmatic meanings" (2011: 382). Armstrong et al. (2018) examined how 3-5-year-old Catalan children understand belief states using auditory and facial prosodic information, and found that children were better at understanding prosodic meanings as they grew older; also, they did better with audio visual prosodic information than just audio alone or visual alone. Work by Hübscher et al. (2017) and Hübscher, Wagner, and Prieto (2020) reported that children as young as 3 years old use both auditory and visual prosody to understand sociopragmatic meanings. They recognize politeness from prosodic or visual cues even before they have fully acquired various lexical and morphosyntactic politeness markers. Shochi, Erickson, et al. (2009) looking at Japanese 9- and 10-year-old children, reported that facial cues helped in understanding politeness and impoliteness expressions, and that the visual cues were perhaps more important for children than for adults.

Long (2010) looking only at the auditory component of prosody, reported that children as young as the first grade begin to be aware of the difference between expressions of apology and gratitude, and by the 7th or 9th grades, their knowledge is similar to that of adults. These findings suggest that language socialization is a process which takes place in the first seven to nine years of school.

For an interesting description of how young children in Japan are taught the general concepts of politeness, see Burdelski (2010): he shows how e.g., enacting role-plays and training social routines (request, acceptation, offer, invitations) help children develop positive, other-oriented behaviors that correspond to Japanese social rules. We are not aware of any studies about when children learn about different categories of politeness; however, generally, around the 5th grade, they start to learn some of the lexical aspects of politeness (Harada 2011).

Our current study, in contrast to our previous studies, examines the multimodal aspect of im/politeness expressions in Japanese society, as it relates to the acquisition of these expressions by children from 6 to 9 years old. Specifically, the current paper proposes a new acoustic and visual analysis of the stimuli, based on two new perceptual experiments we report here.

The im/politeness expressions we examine are the Japanese concepts of teinei, seii, kyoshuku, heijo, and zonzai (broadly translated as 'polite', 'sincere', 'walkingon-eggs', 'declaration', and 'impolite'). These particular expressions were chosen because they represent various types of im/politeness expressions in Japanese. These translations are rough, and cannot be understood without a brief introduction to the hierarchical social structure of Japanese culture, which largely differs from western categories. (The reader is referred to e.g., Long 2010; Ofuka et al. 2000 and references therein for descriptions of the relationship between Japanese culture and sociopragmatic expressions.) In contrast to many western societies where the conversational "other" is often defined in terms of "social intimacy", in Japanese society, the conversational "other", the one to whom the conversation is directed, is culturally defined in terms of their hierarchical relation (e.g., age, social status, power) to the speaker. Social intimacy can play a role, but usually in the use of impolite expressions, discussed below. In general, in Japan the addressee is either at the same level as or below the speaker in terms of age and/ or social status and power, or above the speaker's level. The social situation that is measuring the distance (e.g., school, family, business, etc.) also plays an important role: older sibling/younger sibling, older classmate/younger or same age classmate, more experienced or older colleague in the workplace/ less experienced or younger colleague, etc. Given this importance of social hierarchy, expressions of im/politeness are highly formulaic (Ide 1989). Within the framework of Japanese society, the expressions are all formulaic codes pertaining to how to maintain "wa" (和, 'harmony, peace'), in various social situations; that is, each type of "politeness" is a formulaic expression with a specific name and a specific prosodic expressivity, in addition to a grammatical structure.

The term *teinei* (丁寧) politeness, is used for polite behavior with people like bus drivers, shop clerks, etc. – people who are not well-known, to show respect for a person who is a relative stranger. It can also be used casually with friends. The literal meaning of *teinei* is 'to pay attention' to how to speak; it is the generic term for polite speech which we may translate as 'courtesy politeness'.

Seii (誠意) politeness, deals with how to express politeness in difficult situations with someone either higher than you, or the same level; the focus is on how to enthusiastically respect the sincere feelings of the addressee. We may translate this type of politeness as 'sincere politeness'. The literal meaning of seii (誠意) is 'honest, sincere feeling'. An example of using seii might be expressing respect to another person, even though you disagree with that person. Ruth Benedict, in her classic work "The Chrysanthemum and the Sword", describes a situation in which seii was violated (see Benedict 1947: 159-160). A young man tells an older couple that he wants to go to the United States. They reply with a certain amount of incredulity, which was interpreted by the young man as indicating disrespect to his sincere feelings and the result was the speaker cut off relations with them because their rudeness was unforgiveable. Benedict, in her explanation of the speaker's reactions, points out that sneering, or sarcasm in Japanese culture, is a "wounding of the heart", a type of impoliteness that is difficult, if not impossible, to forgive. This anecdote about seii illustrates the extent to which respect of the other's feelings is extremely important; violation of another's feelings verges on immorality. It also helps the third author of this paper understand better why her "playful" sarcasm was not tolerated by her Japanese colleagues and friends, often leading to severe breakdowns in the relationship, which took a lot of work to mend.

*Kyoshuku* (恐縮) has the literal meaning of 'shrink fear': the speaker "shrinks himself" to show proper "fear/respect" toward his/her superior. *Kyoshuku* is a formulaic politeness strategy concerning how to handle difficult situations with a superior, such as a boss, or a school principal, situations in which showing respect to the superior and his/her point of view is of utmost priority. *Kyoshuku* politeness conveys the speaker's ashamedness / embarrassment that they are not able to act in accordance with the expectations of the person who is hierarchically superior to them (Sadanobu 2004). For example, if a professor has to tell his superior, the school officer, that s/he cannot attend the weekend-scheduled entrance exams due to a personal prior commitment, or, if an employee has to tell their boss that the boss has made a mistake in decision-making, or if a shop clerk has to apologize to the customer for something, regardless of whether it was the shop clerk's fault or not, etc., *kyoshuku* politeness is used. In each case, the priority of the speaker is to respect the superior person's feelings. By using *kyoshuku* politeness, this goal is achieved; it avoids a potential conflicting situation, which may at times

result in getting a favor from the hierarchically superior person. Although young children are never in adult-type boss-employee situations, still they encounter situations where they need to politely reject a *senpai*'s (older classmate) opinion in a way to express respect for the *senpai*'s feelings (see Ide 2002, for examples of hierarchical honorifics used by school-aged children). At what age children begin to use *kyoshuku* is one of the topics of investigation of this paper. It is during the 5th school grade that the lexical items related to honorific form (尊敬語 *sonkeigo*), humble form (謙譲語 *kenjogo*), and courtesy politeness (丁寧 *teinei*) are formally taught in the national education program. The prosodic characteristics of *kyoshuku* politeness, as well as the other types of politeness, are discussed toward the end of this introduction.

Note that for *kyoshuku*, as well as for *seii*, the expressions have formulaic grammatical and prosodic characteristics, independent of how the speaker truly feels about a particular person or fact, thus further underlying the culturally-prescribed formulae for Japanese expressions of im/politeness. Also note that both *kyoshuku* and *seii* perhaps require more cultural awareness of respect for position and age than does *teinei*, which is more similar to western concepts of politeness. As far as we know, there is little research on *seii*; for *kyoshuku*, see Sadanobu (2004) and some of our previous studies, e.g., Shochi (2008); Shochi et al. (2008). *Kyoshuku*, which has no lexical equivalent in English, has roughly been translated as 'walking on eggs' (e.g., Rilliard et al. 2017), because *walking on eggs* is the nearest available in the western inventory of im/politeness, in that the speaker tries to avoid hurting the other's feelings, often of a person who is somewhat difficult to get along with. However, *kyoshuku*, as explained above is clearly very different from *walking on eggs*.

*Heijo* ( $\mathbb{T}$   $\mathbb{R}$ ) politeness is used when the focus is on the facts, not the relationship of the speaker to the addressee. The literal meaning of *heijo* is 'plain sentence'; we may translate it as 'neutral politeness', or 'declaration'. This type of politeness presents information or facts to the hearer without any affective meaning.

Zonzai ( $\mathcal{Z} \wedge \mathcal{Z} \vee$ ) can be roughly translated as 'rude', but has complexities. It is normatively spoken to someone with the same or lower social level. Literally, *zonzai* means 'rude, careless, sloppy', and refers to speech in which the speaker is disrespectful to the addressee. *Zonzai* may be expressed in two ways: first when the speaker voluntarily expresses his "rude" and "arrogant" attitude to the hearer; that is the case of *zonzai* in the stimuli we examine. It can also be expressed by the absence (lack) of polite attitude, and the overt display that the hearer is not part of the speaker's group and is thus a strategy of positive impoliteness (damaging the hearer's positive face; Culpeper 1996). Since age is an important factor to determine hierarchy in East Asian cultures, a *senpai* (older) person has a right (greater possibility) to behave with *zonzai* to a *kohai* (younger) person. Thus, *zonzai* can

© 2023. John Benjamins Publishing Company All rights reserved be spoken to someone who is an equal or a younger person to indicate that "I am *senpai* (older); you are *kohai* (younger)". For instance, in elementary school, a *senpai* fifth grader might speak to a *kohai* fourth grader with *zonzai*. Another example is that sometimes at an after-business drinking party, a *kohai* might drink too much, thus risking behaving with *zonzai* with a higher person; however, in this situation, the *zonzai* behavior often tends to be tolerated. Interestingly, *zonzai* can be used in close relationships, probably mostly between members of the same sex and social status, as a way to indicate intimacy and bonding, frequently seen in current popular anime.

Listing five expressions of politeness in the above order might suggest a linearity of politeness in Japanese culture, from most polite to least polite; however, this is misleading. Expressions of politeness in Japan are not linear, as will be elaborated upon in the course of this paper.

With regard to prosodic features, studies show that social affects are signaled by various acoustic cues, such as fundamental frequency ( $F_{o}$ ) characteristics (intonation contour,  $F_{o}$  height, and amount of change in  $F_{o}$ ), loudness, duration and voice quality (Scherer 2001; Rilliard et al. 2009) and listeners' perceptual patterns are influenced by the gender and the maturity (age) of the listener (Roseberry-McKibbin and Brice 1999).

With regard to acoustic cues of politeness in Japanese, numerous studies (e.g., Kawano 1995; Hong 1993; Ofuka et al. 2000; Ito 2002; Shochi 2008) have reported that politeness is signaled by increased  $F_0$  as well as increased duration, especially on the final vowel (e.g., Ofuka et al. 2000). However, Sherr-Ziarko (2019), in his examination of formality, politeness and  $F_0$ , found that  $F_0$  tends to be higher in informal Japanese speech, thus suggesting  $F_0$  might be viewed as "multi-faceted" issue (2019: 348). As for impoliteness, the final syllable is also lower and longer (Ofuka et al. 2000), and often, stressed alveolar trill [r] can be heard (Calvetti 2020). The prosodic cues of *kyoshuku* are a tense, harsh voice, accompanied by a marked head lowering (Sadanobu 2004), along with a short duration and flat  $F_0$  (Shochi 2008). To date, no research that we know of has been done with *seii*.

Work by Shochi et al. (2008); Rilliard et al. (2009) showed that Japanese adult listeners are able to perceive social affective expressions based on auditory or visual prosody, as well as a combination of both; specifically, L1 listeners were able to recognize a set of twelve Japanese attitudes, while American English and French listeners, without knowledge of the Japanese language, showed important confusions between some expressions. More specifically, the *kyoshuku* expression was not categorized within the polite expressions by non-native listeners, but linked to irritation or *zonzai*, while the other politeness expressions received high identification scores (Shochi 2008). One motivation of our current work is thus to investigate the interrelations between expressions of politeness in Japanese, in order to measure if and when Japanese children start to identify social im/polite expressions and link them to adequate social contexts as adults do.

The current paper proposes a re-analysis of a perceptual judgement experiment with children (Shochi, Erickson, et al. 2009), in the light of new results from a study on pair comparisons with young (6–9 yo) children. It proposes a new analysis of these two sets of perceptual results in line with a new acoustic and visual analysis of the stimuli. Section 2 presents the stimuli used and its analysis, and the listeners involved in the test as well as the experimental paradigms. Section 3 describes the results, and Section 4 gives an analysis of the main findings and their implications for the strategies of politeness in Japanese.

#### 2. Methods

#### 2.1 Listeners

The listeners all spoke Kumamoto dialect. Given that the standard Tokyo dialect is prevalent in national television programs, most listeners had familiarity with the standard Tokyo dialect. Concerning the first experiment, three groups of listeners participated: The first group consisted of 29 adults (18 females, 11 males; mean age=23); the second, 28 children in 4th grade (G4) in the Japanese school system (13 females, 15 males; mean age=9.9 years); the third, 50 children in 5th grade (G5) (31 females, 19 males; mean age=10.9 years). The total number of participants for this experiment was 107.

In the second experiment, four age groups participated, with all listeners different from the first experiment. The first group was composed of 18 children in 1st Grade (11 females, 7 males; mean age: 6.1), the second, 19 children in 2nd Grade (13 females, 6 males; mean age: 7.4), the third, 19 children in 4th Grade (9 females, 10 males; mean age: 9.5), and the fourth, 40 adults (28 females, 12 males; mean age: 21.6). The total number of participants was 96. The explanation for the different age groups in experiment 2, compared to experiment 1, is as follows: The aim of experiment 2 was two-fold: to examine children of a larger set of ages and to compare the results with experiment 1. Experiment 1, done prior to experiment 2, only examined 4th and 5th graders, since the questionnaire task in experiment 1 was too difficult for younger students.

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. No listener reported any hearing problems. 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. In accordance with the ethical rules of each of the elementary schools, the parents signed a consent form before the tests allowing their children to participate in the experiments.

# 2.2 Stimuli

Five Japanese expressions related to politeness were examined: *teinei* (TEI), *seii* (SEI), *kyoshuku* (KYO), *heijo* (HEI) and impolite *zonzai* (ZON). Each audiovisual stimulus was recorded using the same affectively neutral sentence as shown in (1) to express the prosodic attitudes (Shochi, Rilliard, et al. 2009) by a trained Japanese language teacher, the first author of this paper, who has considerable experience teaching social affective expressions to learners of Japanese.

 (1) 名古屋で飲みます Nagoya denomimas Nagoya have-a-drink
 'Let's have a drink in Nagoya'

Given that polite expressions in Japan are highly formulaic, the recordings represent exemplar productions of the five types of politeness following the definition and the context inspired by previous research (Sadanobu 2004; Kawano 1995; Hong 1993; Ohara 2001; Ofuka et al. 2000; etc.). The five polite expressions were validated previously by adult Tokyo dialect listeners (Rilliard et al. 2009; Shochi et al. 2008). The topic of gender variation in the production of these types of politeness is a topic for future exploration.

The recordings were done in a soundproof room at LIMSI, France using a digital DV camera (Canon XM1 3CCD) and an omnidirectional AKG C414B microphone placed 40 cm from the speaker's mouth. The distance to the microphone and the recording level were kept constant, thus allowing for reliable intensity measurements (i.e., comparable between sentences) if the absolute level (SPL) is unknown. The microphone was connected to a USBPresound device connected to a computer outside the room, recording the speech signal at 44.1 kHz, 16 bits. In order to replace the camera sound with the high-quality sound recorded by the microphone, synchronization was achieved with hand claps performed between each sentence. Video clips were encoded with a cinepack codec with a 784 × 576 pixels resolution.

Each stimulus was presented to the elementary school listeners either with audio-only, video-only or audio-video modalities. The acoustic and facial movements typical of each of the five expressions are described in the next section.

#### 2.3 Acoustic and facial analysis

#### 2.3.1 Acoustic and facial measurements

The following acoustic parameters were extracted from the five utterances of Nagoya de nomimasu, spoken with teinei, seii, kyoshuku, hyojo, and zonzai, using Praat (Boersma and Weenink 2020): (i) the voice's fundamental frequency  $(F_{0})$ measured in semitones (with reference to 1 hertz), (ii) the signal intensity (in dB), and (iii) the moraic durations (in seconds). Then, the following visual parameters were estimated from the video thanks to the OpenFace software (Baltrušaitis et al. 2018; Baltrušaitis, Mahmoud, and Robinson 2015): (iv) the amplitude (on a 1 to 4 scale) of a set of Action Units (AU) (Ekman, Friesen, and Hager 2002) from the upper face (AU01, 02, 04, 05, 06, 07, 09 & 45), and (v) a set of AU from the lower face (AU10, 12, 14, 15, 17, 20, 23; AU on lip opening and jaw drop were not considered because they were biased by articulation). An overview of the AUs can be found in Table 1. Then, (vi) the rotations of the head were measured over three axes (pitch, yaw, and roll); and (vii) gaze direction was estimated along x- and yaxes (i.e., x: left/right movements, with negative values indicating right gaze; y: up/down movements, with negative value indicating up gaze). Left and right are given from the position of the speaker.

Action Unit	Description	
AU01	Inner brow raiser	
AU02	Outer brow raiser	
AU04	Brow lowerer	
AU05	Upper lid raiser	
AU06	Cheek raiser	
AU07	Lid tightener	
AU09	Nose wrinkler	
AU10	Upper lip raiser	
AU12	Lip corner puller	
AU14	Dimpler	
AU15	Lip corner depressor	
AU17	Chin raiser	
AU20	Lip stretcher	
AU23	Lip tightener	
AU45	Blink	

**Table 1.** Description of the action units used in this study, as found in (Ekman, Friesen,and Hager 2002)

The detail of the facial expressions is displayed through a series of still images extracted from each video, and presented in Figure 1.



**Figure 1.** Each row features a series of six still images extracted regularly from the video of each social affect and cropped so to focus on the head; the expressions are presented in the following order, from the top: *teinei, seii, kyoshuku, heijo, zonzai* 

These audio-visual measurements were taken each 5 ms for  $F_0$  and intensity, for each mora for moraic duration, and at each video frame (one each 40 ms) for visual parameters; they were aligned and linearly interpolated in order to get vectors with a 5 ms time frame (except for duration). These values were then regrouped (taking their mean value) for each mora of the sentence to create the figures displayed in the next section that present the mean and confidence intervals for the values observed on one mora.

# 2.3.2 Acoustic parameters variation

 $F_{o}$  and intensity levels (Figure 2) tend to mark the difference between attitudes, with louder performance, but not higher  $F_{o}$ , for *zonzai*. *Kyoshuku* shows the flattest curve, while on the contrary the *teinei*, *seii* and *heijo* expressions show large pitch spans (especially *teinei*); *seii* has a relatively high pitch for a low intensity performance. As for the interaction between  $F_{o}$  and intensity, we see that for the two expressions of politeness, *seii* and *teinei*,  $F_{o}$  increases while intensity decreases. The other three expressions, *heijo*, *kyoshuku* and *zonzai*, suggest that  $F_{o}$  and intensity work together. The pattern of high  $F_{o}$  and low intensity found for the two polite expressions can be seen as characteristics of low dominance utterances, and opposed to the strategy of power display used for *zonzai* with higher intensity and relatively low pitch, along the lines of the Frequency code (Ohala 1983, 1994) – with *teinei* also using a large pitch span typical of the Effort Code's predictions (Gussenhoven 2004). On the contrary, *kyoshuku* did not fit any of the predictions of these two symbolic codes, as will be discussed later.



**Figure 2.** Plot of the mora's means and confidence intervals for  $F_0$  (top plot), intensity (middle plot), moraic duration (bottom), for each attitude (left to right graphs). Duration values are raw values for each mora

As for duration, the final mora /su/ shows the longest duration in the *kyoshuku* expression. In general, however, the moras with the open vowel /a/ (i.e., /ma/, /na/) have the longest durations, with the /ma/ in *zonzai* being the longest. Lengthening of the penultimate mora for *zonzai* is a prosodic characteristic of impoliteness expressions (see e.g., Ofuka et al. 2000; Calvetti 2020).

# 2.3.3 Action Units

Only AU movements higher than an intensity of one are considered here: under that rule, most action units are silent for these expressions for this speaker, except for the *kyoshuku* expression (see Figure 3). In the case of *kyoshuku*, the AUs 04 (brow lowerer), 07 (lid tightener), 10 (upper lip raiser), and 17 (chin raiser) are activated during the sentence, forming a frowned face; slight markings of AUs 12 (lip corner puller), 20 (lip stretcher), and 23 (lip tightener) are also used (not shown) and participate in the global contraction of the face.



**Figure 3.** Plot of the mora's means and confidence intervals for AUs 04, 07, 10 and 17 for each attitude (left to right graphs)

#### 2.3.4 Head movements

The rotations of the head are shown in Figure 4 for each expression. The main movements are observed for (i) *kyoshuku*, with the head going down throughout the sentence, and turned toward the right (with a slight roll on the left that may result from the other movements); (ii) *zonzai*, with head up at the beginning, turned toward the left during the sentence (movement opposite to that of *kyoshuku*); (iii) *teinei* and *seii*, with slight head down at the end of the sentence (larger movement for *teinei*). Notice that *teinei* shows a wider  $F_0$  span than *seii*, suggesting a possible connection between  $F_0$  and head position. The head is mostly still for *heijo*.



**Figure 4.** Plot of the mora means for head pitch (top, positive value for head down), yaw (middle, positive for right movement), and roll (bottom, positive for left), for each attitude (left to right graphs)

# 2.3.5 Gaze direction

Gaze directions are shown in Figure 5. During the *zonzai* expression, the speaker is staring at the addressee while his head is up, which makes it a typical "looking down" type of pose (head up, with gaze directed at the interlocutor). For *kyoshuku*, the gaze is down throughout the sentence, while for *teinei* and *seii*, the gaze goes down to a lesser extent, especially at the end of the sentence. Gaze also goes rapidly down in the middle of the sentence for *heijo*. For *heijo*, there is a rapid fall of  $F_0$  on the third mora of the utterance. Notice that head and gaze movements tend to synchronize.



**Figure 5.** Plot of the mora means for the two gaze directions, red for horizontal (positive for left) and blue for vertical positive for down), for each attitude (left to right graphs)

Table 2 presents a summary of the main facial and vocal changes found in each expression, basing the description on the *heijo* expression as the most neutral one, and describing the others in relation to it.

Measure	Teinei	Seii	Kyoshuku	Heijo	Zonzai
F <sub>o</sub>	highest start; lowest end: larger range	initial equal to <i>heijo</i> , fall down to the mean (not lower)	flat; close to mean	high initial, falling below the mean for a low flat end	Relatively low (given the high vocal effort), decreasing F <sub>o</sub>
Intensity	peak on /no/	peak on /mi/	peak on /mi/; fall after	above the mean for /nagoya/; fall on the 2nd word (see $F_0$ )	strongest initial, with important decrease on the second word

 Table 2.
 Summary table of the main characteristics of each expression. Note that for AUs, the movements that do not go above an intensity of 1 were not considered here

Measure	Teinei	Seii	Kyoshuku	Heijo	Zonzai
Lengthening	long /masu/	long /ma/	long /masu/	long /ma/ strict isochrony on the other	long /ma/ important lengthening (twice heijo duration)
Action units	0	0	AU04+07: brow frown AU10+17: frowned face	0	0
Head pitch	++ at sentence end	+ at sentence end	+++ along the sentence (bowing)	0	- (start)
Head yaw	0	0	Turn right	0	Turn left
Head roll	0	0	Tilt left	0	Tilt left
Gaze horizontal	o (addressee direction)	0	Left (indirect)	0	0
Gaze vertical	down at end	down at end	down	rapid down in the middle	straight

Table 2.	(continued)
14010 21	(001111110000)

# 2.4 Experimental paradigm

Two experimental methods for analyzing the perception of five different stimuli being labeled as "*teinei, seii, kyoshuku, heijo*, and *zonzai*" were carried out: (i) the perceived degree of politeness in terms of the category of people who may be addressed in such a way by the speaker, and (ii) the similarity between pairs of expressions.

For the first experiment, the test consisted of 15 different stimuli (5 attitudes \* 3 modalities), each presented three times to listeners in a random order. For the second experiment, the same five stimuli were used in a pair-comparison paradigm, where participants had to rate the similarity between two expressions, in terms of the comparability of speaker behavior; ratings were given on a 1 to 9 scale. All pairs of the five expressions were presented in audio-only, visual-only, and audio-visual modalities.

# 2.4.1 Experiment 1

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). In order to balance the effect of the presentation order, half the subjects began with the video-only stimuli and the other half with the audio-only ones. Each stimulus could be listened to only once.

After each 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 could easily understand. They were then asked to judge the degree of politeness (*teinei*) of the stimulus on a scale ranging from impolite to polite (encoded on a 0–10 scale); the middle of the scale corresponding to a neutral expression. The listeners were asked to use a slide bar displayed on the computer screen to move a marker to indicate their judgments; the position of the marker was set to the middle of the scale for each presentation, and subjects had to move it according to their perception of the degree of politeness.

Each subject participated in 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.

### 2.4.2 Experiment 2

Participants were asked to judge the difference between pairs of stimuli, on a 1 to 9 scale (1 for almost identical stimuli, 9 for completely different ones). Stimuli were based on performances of the same sentence presented in Example (1), and produced so as to convey five different prosodic attitudes, *teinei, seii, kyoshuku, heijo*, and *zonzai*. Each possible pair of the five performances (i.e., ten pairs), in each modality (A, V, AV), were created in AB or BA order, separated by a short pause, in order to be presented to participants.

Participants were instructed to give their answers as fast as possible on the basis of their first impression. The AB/BA presentation order, as well as the order of presentation for modalities (either A-V-AV or V-A-AV), was balanced across participants.

#### 2.5 Statistical processing

#### 2.5.1 Experiment 1

Results of 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 a correspondence analysis (CA; Husson, Lê, and Pagès 2017). 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" (Clm) and "unknown" (Unkn) – for each presented attitude, the proximity of attitudes and of social status was described. Different analyses were carried out for each of the three groups of judges (adults, G4 and G5), mixing answers of the three modalities.

The answers to the second question – the degree of politeness – were analyzed using the 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 repetitions of each stimulus (R – 3 levels, fixed factor). Two between-subject factors were used: age of subjects/level in school (G4, G5, or adult) (L – 3 levels, fixed) and the order of presentation of modalities during the test (O – 2 levels, fixed). Following Max and Onghena (1999), the Huynh-Feldt correction was applied to all results of the repeated measures ANOVA.

#### 2.5.2 Experiment 2

The analysis of the second experiment follows the methodology presented in the work by Romney and colleagues (Romney et al. 1996; Romney, Moore, and Rusch 1997; Romney et al. 2000): it consists of applying a correspondence analysis to the similarity matrix obtained from the pair comparison results, in order to obtain a spread of the presented stimuli on a multidimensional space. This spread of stimuli is seen as representing the structure of perceptual proximity between the stimuli. Similarities and differences in this spread of stimuli across groups is then quantified to compare their coherence in terms of "shared knowledge" (Moore et al. 1999).

The next step consists in studying the spread of stimuli across the perceptual space. First, for each pair of stimuli composed of two attitudes A and B, their similarity score was obtained from subjects' answers: distance estimations obtained from subjects were converted into similarity scores, with a 10 for pairs of identical attitudes (AA pairs, not presented to participants), and a 10 minus the measured distances for the heterogeneous pairs (the distance being on a 1 to 9 scale); this was done because correspondence analysis is designed to work on similarity data (cf. Romney et al. 2000). Then, these similarity scores were stacked in a 5×5 similarity matrix (experiments are based on five attitudes: each line of the matrix contains the similarity of a given attitude vis-à-vis the set of the five possible attitudes). One such 5×5 matrices were stacked in a 1440 (5 attitudes × 3 modalities × 96 subjects) × 5 matrix, hereafter referred to as X. A Correspondence Analysis (CA) was performed on X (Romney et al. 2000). The row scores obtained form the CA were standardized (Kumbasar, Romney, and Batchelder 1994), so that for

each subject, the values are set to a mean of o and a variance equal to the singular value of each dimension of the CA. This standardization allows scaling of the data for possible differences in the use of the answer scale between subjects. The effect of such a normalization may at worst reduce the strength of the observed variations, according to Romney et al. (1996). The CA on matrix X gives a representation of the perceived distances between the 5 attitudes, for each subject and in each modality. Raw results give a cloud of points where the main tendencies can already be seen. By grouping scores from subjects of the same grade and from the same modality, it is possible to obtain a figure of the effects of age and modality on the representation of these 5 attitudes.

The output of the CA has 4 dimensions, but the first two factors for the row account respectively for 38.7% and 34% of the total variance: i.e., a 2D plot shows 72.8% of the variance in the distance between attitudes. In the corresponding plots (see the result section), all individuals will not be represented; rather, it shows the mean coordinates on the first two factors for each category observed (with the ellipses around the means representing the 97.5% confidence limit from the means). The categories used to group answers were: (i) the three presentation modalities (A, V, AV) and (ii) the four age groups (1st grade, 2nd grade, 4th grade and Adults).

We then look at the shapes of the perceptual spaces. From the spread of stimuli in the CA space, the shapes of the "perceptual spaces" obtained, thanks to the spread of points on the CA space, are compared across groups, using the method presented in Romney et al. (2000), inspired by Rao and Suryawanshi (1996). The comparison is based on the measurement of the Euclidean distances between all pairs of points representing attitudes in the CA space (the four dimensions are used). The obtained distances give a description of the shape of the five points in the CA space, which is supposed to reflect the "perceptual space" defined for these five attitudes and for each participant in each modality. To compare the perception of the different participants, the correlations between the vectors describing these perceptual spaces are calculated for each pair of participants in each modality. The result is a 288×288 correlation matrix (3 modalities × 96 subjects), on which a Principal Component Analysis (PCA) is run. A plot of the mean position of the point by age group or by modality groups allows the observation of changes in shapes linked to these two factors.

Finally, the variance in the results is used to estimate the proportion of "shared knowledge" across participants. If one takes the square root of these correlation vectors (i.e., the correlation between the shapes of the five-point spread obtained for each participant), it allows an estimation of what was coined by Romney et al. (2000) as the knowledge shared by the different groups. The principle is that the more correlated the two "perceptual spaces" are, the more the two participants

share similar cultural knowledge for the data under investigation. Values reported can be viewed as the ratio of shared knowledge between different groups of participants. A high within- or between-group ratio demonstrates their coherence in the rating task; conversely, a low score demonstrates large variation in responding to the task.

#### 3. Results

#### 3.1 Experiment 1

#### **3.1.1** *Question* 1

Results for question 1, about who may be addressed in such a way by the speaker, are summarized in Figure 6. The figure shows the spread of attitudes and potential addressees along the two first dimensions (that represent about 90% of the variances) of the CA run for each group. The main similarity across age groups is the systematic association by all subjects of the answer "Lower grade student/person" to the *zonzai* expression on one side of the first dimension, while the other four expressions are on the other side of this axis.

Amongst these polite expressions, adults associate *kyoshuku* to "Higher grade student/person", *teinei* and *seii* to "Teacher", while *heijo* is used with "Unknown" addressee. No systematic associations are made with "Classmate". Both G4 and G5 children tend to associate *teinei* and *seii* with "Teacher"; G4 associate *kyoshuku* with "Classmate", while G5 with "Higher Grade student", in a fashion similar to adults. These results suggest that G5 students are approaching an adult-like understanding of *kyoshuku*, but G4 students still do not have a complete grasp of the social use of *kyoshuku*.

#### 3.1.2 Question 2

Figure 7 presents the mean degree of "politeness" given by the three test groups to the attitudes, in each presentation modality.

Diagrams are used in Figure 7 as a way to compare the similarities and differences among the age groups and the modalities. In terms of overall differences in levels of politeness, we see that *teinei* is clearly more polite than *heijo*, with *zonzai* clearly least polite. So, on one hand we see a gradual negative progression of politeness levels. On the other hand, we see that *seii* and *kyoshuku* are perceived in all modalities and by all age groups as less polite than *teinei*, indicating that these two types of politeness fall outside the polite-impolite dichotomy, thus, underlining the non-linear, multidimensional character of politeness expressions in Japanese culture.

© 2023. John Benjamins Publishing Company All rights reserved



**Figure 6.** Position of the stimuli (blue points) and of the addressees (red points) presented to each age group (Adult, G<sub>5</sub>, G<sub>4</sub>) along the first two dimensions of the CA (see text)

© 2023. Jonn Benjamins Publishing Company All rights reserved



**Figure 7.** Mean and confidence interval of politeness levels attributed to the five expressions (x-axis) by each group of judges (individual lines), in each modality (individual plots)

For the audio modality, we see for all three groups a progressive increase in ratings of politeness: *zonzai*<*kyoshuku*<*heijo*<*seii*<*teinei*, except that G4 children perceived *seii* as slightly less polite than *heijo*. Notice that G4 children rate *zonzai* as more polite than the other groups, suggesting that perhaps they are not aware of some of the acoustic prosodic features of impoliteness. It is interesting that in the audio modality, all three groups rate *kyoshuku* with approximately equal amounts of politeness. Notice also that children in G4 and G5 perceive *teinei* as more polite than adults do.

For the video modality, we see basically the same order of increased perception of politeness. But there are some differences. Now, all three groups equally rank *teinei* with the same degree of being most polite. Adult listeners perceive more "politeness" (*teinei*) for *kyoshuku* with the video presentation compared to the audio presentation; however, G4 and G5 children perceive less politeness for *kyoshuku*. This suggests that children are not yet attuned to the visual prosody of *kyoshuku* familiar to adults. It is interesting as well as puzzling that G4 and G5 children perceive *heijo* as being more polite than adults in the video presentation.

For the audio-visual modality, we start to see more similar ratings of *zonzai* impoliteness for the three groups, with G4 perceptions moving closer to those of G5 and adults. As for *kyoshuku*, G5 perceptions approximate more closely those of adults. An interesting and somewhat puzzling difference is that G4 and G5 children tend to rate *heijo, seii* and *teinei* as more polite than adults.

According to the ANOVA, the factor "attitudes" had a main effect on listeners' answers, as well as the factor "modality of presentation", and the interaction between both factors. Moreover, the interaction between attitudes and the group

of subjects (G4, G5, or adults) was also significant. The interaction between attitude, repetition and the order of presentation was significant, but seems difficult to interpret. As for the effect of the attitude factor, results showed that the main effect of attitude on the perceived degree of politeness puts zonzai at the lowest end (the impolite end) of the answer scale (mean = 2.7), while all the other expressions received scores over 5. Teinei was rated as the most polite (6.7), with seii receiving the next highest score (6.1). Heijo received a mean score of 5.8, while kyoshuku got a 5. The interaction attitude\*level is represented in Figure 7. Interestingly, the politeness degree of kyoshuku systematically rose with age, from 4.7 at G4 and 5 at G5 to 5.3 for adults, while it decreased for *heijo* (G4=6.3;  $G_5 = 5.9$ ; AD = 5.3). Zonzai also received lower scores from older subjects ( $G_4 = 3.3$ ; G5=2.4; AD=2.6), while seii and teinei didn't change much with age. As for the effect of the modality factor, 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 zonzai. The interaction attitude\*modality was significant ( $F_{(8,804,4)}$ =8.19, p<0.000), and with the audiovisual presentations, listeners use the widest range of politeness ratings: politeness expressions get the largest contrast with the impoliteness expression, zonzai. However, the patterns of answers in each modality were similar across attitudes with the exception of zonzai, for which the visual information helped listeners to perceive the impolite expression (A=3.4; V=2.5; AV=2.1).

#### 3.2 Experiment 2

#### 3.2.1 Spread of stimuli across the perceptual space

Figures 8, 9, and 10 present the distribution of the mean position of the five attitudes (shape of points) observed for each age group (colors), respectively for each of the three presentation modalities.

The global shape of the spread of the five attitudes is stable across grades and modalities. *Zonzai* and *kyoshuku* are the expressions that are the most separated from others; the differences between *heijo, teinei* and *seii* vary depending on the presentation modality, with more separation observed in the audio-only presentation, while the visual presentation mostly increases the distance of *kyoshuku* to the *teinei/seii* group along the vertical axis. Age group differences are mostly visible for the younger children (1st grade), who showed a reduced space spread in the visual-only presentation; generally, the audio-visual presentations allowed a clearer separation between age groups.



**Figure 8.** Positions of the five attitudes (point shapes) for the four age groups (colors), averaged over results obtained for all participants for the audio-only presentations; ellipses present the 97.5 confidence intervals



**Figure 9.** Positions of the five attitudes (point shapes) for the four age groups (colors), averaged over results obtained for all participants for the video-only presentations; ellipses present the 97.5 confidence intervals



**Figure 10.** Positions of the five attitudes (point shapes) for the four age groups (colors), averaged over results obtained for all participants for the audio-visual presentations; ellipses present the 97.5 confidence intervals

# 3.2.2 Shape of the perceptual space

Figure 11 presents the spread of the mean difference between the shapes of stimuli spread (the shapes of the perceptual organization of the five attitudes in the CA space; what Romney et al. 1996 call the "cognitive representation" of these expressions), averaged over age groups (color) or across presentation modality (grey). The result of the PCA on the correlation between the shapes of the five attitudes' spreads shows a progression of shape linked to age, from the lower left corner for 1st grade children, to the upper right quadrant for adults.

In Figure 11, we see the effect of the modality of presentation on the shape of stimuli spread, with audio cognitive representations (A) being in the upper left; visual representations (V) in the lower right quadrant and AV representations in the middle. The effect of age group is shown in terms of colored ellipses around the means, indicating a consistent progression of the shapes from 1st grade up to the adult organization of these stimuli. Notice that the position of the adult data corresponds to the position of audio-visual shapes, which appears to be a mix of the dispersion obtained in the audio-only modality (upper-left quadrant) and the visual-only modality (lower-right quadrant); for the children, we see a gradual integration of audio and visual modality with increasing age, and awareness of cultural norms of politeness.



**Figure 11.** Spread of the perceptual shapes along the first two dimensions of the PCA run on the correlation matrix; the points indicate the mean position for each age group (colored points and ellipses), and for each presentation modality (grey points and ellipses)

# 3.2.3 Proportion of shared knowledge

The mean ratio of "shared knowledge" between or within age groups is presented in Table 3. Within group ratios indicate the internal consistency of answers between participants of the same group (found along the diagonal of Table 3); to say it in terms of Moore et al. (1999), these ratios correspond to intra-group shared cultural knowledge.

	1st	2nd	4th	Adults
1st	.55	•44	•45	.60
2nd	•44	.58	.55	.59
4th	•45	•55	.63	.65
Adults	.60	.59	.65	.71

**Table 3.** Square root of the average correlation of shapes, between (off-diagonal) or within (along the diagonal) age groups of participants

Table 3 compares participants from the same culture, but having a growing knowledge of this culture with age (differing in that respect from Moore et al. 1999, who compared groups of adults with differing cultural backgrounds). The adult group, supposedly having the most developed "cultural knowledge", serves as the reference point: results for the child groups are analyzed by comparison to

this reference. The mean Intra-Adults root squared correlation is .71, which indicates the global agreement, or shared cultural knowledge, between judgements of adults on these five stimuli for the three presentation modalities. The remaining 29% of unexplained variance can be attributed to individual differences, sampling and errors.

The decreasingly lower intra-group coherence obtained for groups of children with decreasing age shows that the younger the children, the less they share common knowledge to interpret the expressions (i.e., most of their judgement relies on their own, less standardized, cultural understanding of the expressions). The mean square root of the mean correlation of the three groups of children increases with age, from .55 up to .63 (i.e., by about 3% per year). This may give an idea of the rate at which children learn the norms of their culture with age – and it indicates that, by 4th grade (about 9-years-old), children are already close to the mean adult knowledge (only about 8% lower, a distance corresponding to about three years of learning if they continue at the same mean pace).

The off-diagonal square roots of the mean correlation are, in such an experimental setup, difficult to interpret. They are supposed to reflect the inter-group sharing of common knowledge. If we look at the scores between adults and each group of children, they are similar for 1st and 2nd grade, and increase slightly for 4th grade children. It may be seen as another estimation of the growing cultural knowledge of the children.

# 4. Discussion and conclusions

The results of these two experiments support the idea that the Japanese im/politeness conceptual space has a multidimensional structure. The results of experiment 1 (Figure 7), for which children and adults were asked to rate the politeness of each expression, indicated a somewhat one-dimensional discourtesy / courtesy dimension of politeness, progressing from *zonzai* (impolite) to *kyoshuku* to *heijo* to *seii* to *teinei* (polite), suggesting that *kyoshuku* was in the middle of the scale, closer to *zonzai* than the *teinei* and *seii* expressions in terms of degree of politeness. However, looking at the results of experiment 2 (Figure 10), in which a correspondence analysis was applied to a similarity matrix obtained from the pair comparison results in order to obtain a spread of the presented stimuli on a multidimensional space, we see a different pattern. It shows an opposition between *zonzai* (impolite) and *teinei* (polite) – both being at opposed positions along the first dimension – which reflects this dimension of courtesy politeness. But notice how *kyoshuku* can be seen to *not* fall along this unidimensional line of politeness; rather we see a projection of *kyoshuku* along an orthogonal dimension (almost vertical in Figure 10, that goes down to the bottom right corner-mostly along dimension 2, from a neutral position along the courtesy-discourtesy dimension). The fact that *kyoshuku* does not fall along a unidimensional line from "polite" to "impolite" underscores the complex multidimensionality of the structure of politeness cultural schema in Japanese. These coherent results from the two experiments demonstrate the multidimensional structure of the Japanese im/ politeness conceptual space.

This structuration of the categories of politeness behaviors defined within Japanese culture show that they do not fit on a one-dimensional opposition between impolite impositions (corresponding to zonzai in the case of Japanese culture) on the one hand, and behaviors following courtesy routines on the other (corresponding to teinei). The seii and especially the kyoshuku categories correspond to routines that encapsulate given situations (with a specific hierarchical relation, type of targeted speech acts, etc.), and related accepted behaviors (that include the audio-visual prosodic patterns described here, but also specific formulaic expression and grammatical markers) that do not fit along this courtesy-discourtesy dimension. The courtesy-discourtesy dimension found in the opposition between teinei and zonzai mostly fits the opposition of "politenessimpoliteness" for English (and also at least for most western European cultures). The expression of kyoshuku, however, uses a different strategy, one that fits in with the typology of "negative politeness" as a Face Threatening Act redresser described by Brown and Levinson (1987) (typically within their strategy 6 of negative politeness, that consists of "indicate reluctance"; 1987:188). Also, it fits in with Ide's idea of a high discernment culture (Hill et al. 1986; Ide 2002): it requires the use of culturally defined routines in order to behave according to what is expected for an individual in a given situation. The conceptualization of kyoshuku as a negative politeness strategy is coherent with the results of our second experiment that show this behavior fits along a second dimension of the multidimensional space of "politeness" conceptualized within the Brown and Levinson framework.

Japanese children learn the components of each of these cultural concepts through socialization, but they are also explicitly taught the components of these categories (see Burdelski 2010; Harada 2011). They continue learning these components and their social use as teenagers, as shown by the pace of increased "shared knowledge" within child groups, compared to adults in experiment 2. The vocal strategy linked to *kyoshuku* expresses "suffering" (Sadanobu 2004), using a tense voice and a low flat intonation, an expressive choice that does not fit with expressive codes selected in western strategies of negative politeness (typically the expression of excuses). When presented in its audio-only modality, listeners from American and French cultural backgrounds do not interpret it as some-

thing related to politeness strategies (Shochi et al. 2007; Shochi, Rilliard et al. 2009). But when the expressions are presented in audio-visual modalities, the facial frown associated to a strong bowing gesture disambiguates the vocal quality: the performances are associated to either the highest "politeness" levels by French listeners along a one-dimensional scale (Shochi et al. 2014), or follow a similar patterning of multidimensional distribution (similar to the experiment 2 presented here), when using a pair-comparison protocol presented to Brazilian, American, and French participants (Rilliard et al. 2014). The comparison of our present findings on the acquisition by L1 Japanese children of these expressions to the judgements made by non-Japanese participants on the same productions support the following two ideas.

First, the voice quality used in Japanese society does not correspond to the choice made by American and French cultures: this voice quality, related to the affective expression of suffering, vehiculates a symbolic notion that is certainly captured by the American and French listeners, but that does not match their folk notion of "politeness", conceptualized basically as a strategy of courtesy (i.e., positive politeness). Meanwhile, the audio-visual behavior contains strong cues to polite demeanor (notably the bowing), that disambiguate the use of this voice quality (the vocal cues fit with the visual cues: see Shochi et al. 2020 for another example of audio-visual matching that brings coherent results), and the expressive strength of the kyoshuku behavior even leads American and French participants to rate it as "super polite" (highest "politeness" scores: Shochi et al. 2014) - while it is not judged as especially "teinei" by the Japanese adults. It seems the folk notion of "politeness" thus differs in the two cultures (or set of cultures - Japan vs. France and USA in this case), with Japanese seeing "politeness" as a complex multidimensional set of categories (that includes positive and negative politeness strategies), while the "western" participants rather see "politeness" under its positive politeness conception only - as an expression of courtesy.

Second, if the folk notions of "politeness" in these two sets of cultures differ (and thus their social use of the related categories differ – a fact that is described by Wierzbicka 1985, 1996), the understanding of the symbolic aspects of the speaker's performances does not necessarily diverge to the same extent: this is one of the results obtained with cross-cultural studies. A first study (Rilliard et al. 2014) on the basis of the same experimental protocol as the one used for experiment 2 (pair comparison), shows that the distribution of these five Japanese expressions (*teinei, seii, kyoshuku, heijo, zonzai*) is mostly similar across participants from Japan and non-Japanese speaking participants from Brazil, USA, and France, whatever the presentation modality (audio, visual, or audio-visual). Even the audio-only modality, that carries this specific voice quality, is interpreted in a similar fashion by Japanese and non-Japanese participants. This result shows

the strength of lexical categories and folk cultural conceptualizations affecting the judgements made on our spoken performances. Presenting lexical categories drives opposite results across cultures when asking participants to judge "politeness", while comparable ones are obtained when there are no labels imposed on their judgements. This was illustrated in a second study (Shochi et al. 2020) which asked listeners to associate audio and visual performances in order to recreate a performance that best matches a given label: the use of audio performances of kyoshuku was not selected by non-Japanese (western) participants to match any labels in the sphere of politeness, while it was by L1 Japanese participants. This further strengthens the idea of a conceptual level mismatch between cultures, but not of a performance level mismatch - at least in the case of the expressions investigated here. If performances match cross-culturally, outside linguistic and cultural conceptualizations, this supports the view that they carry symbolic content. In the case of these five expressions, Ohala's Frequency code (1983, 1994) explains the opposition along the line of positive politeness, with low effort, relatively higher pitched teinei and seii, opposed to stronger and lower zonzai. The effort code may explain the relatively large F0 range found in *teinei*, but certainly further analyses of seii and teinei are required to study their similarities and divergences. Kyoshuku, that fits along another dimension that we linked to the set of negative politeness strategies, does not enter this classification, but is related to emotional performances (in this case linked to "suffering": Sadanobu 2004), whose most phylogenetically developed aspects may have been used in order to build more complex expressions or cultural emotions and attitudes (e.g., Scherer 2009) inclusive of kyoshuku.

A second important result concerns the progression of children's understanding of these social expressions. The first experiment observed the progressive building of social hierarchy at the 4th and 5th grades (about 9 to 11 year-olds, see Figure 6); the second experiment expresses this acquisition in terms of the percentage of shared knowledge, through the similarities found in the distribution patterns across judges (Table 3). The two findings are coherent, with 5th graders (exp. 1) already using adult's social hierarchies, while 4th graders (exp. 2) are building coherence as a group, approaching adults in their classification ability but still having about three years to reach it (using a prospective projection of the progression rate of the three age-groups up to adult results).

A third result of these experimental settings concerns the importance of audio and visual modalities in the expression of such social affects. Audio-visual presentation systematically led to more coherent answers, and to larger differentiation capacity across the expressions (Figures 7 and 10), similar to what was reported by, for example, Hübscher et al. (2017); Hübscher, Wagner, and Prieto (2020) and Armstrong et al. (2018). The participants are thus able to gather complementary information from each modality. The complementary aspect of the distribution obtained in the audio-visual modality is clear in Figure 10, with bimodal presentation having shapes in between the two monomodal ones. Comparing audio and visual distributions of the five attitudes (Figures 8 and 9), one may observe the smaller space occupied by the audio answers along the CA dimensions, compared to the visual modality: this reflects less distinctive expressions in the audio compared to the visual modality. Conversely, the five attitudes have separate positions in this smaller space, that are shared across age groups, while the *heijo, teinei* and *seii* expressions are mostly at the same position (the main differences being related to age groups) in the visual modality: this reflects the smallest number of expressive categories expressed through the visual modality when compared to the audio one for this set of expressions. In summary, the visual mode expresses a smaller but more clear-cut set of expressions, while the audio modality expresses various shades of meanings, that are sharing a smaller space.

The expressions of *teinei* and *seii* show little differences in their social use and expressive classification. They participate in a progressive scale of social use that fits the "discernment" part of social interaction proposed by Hill et al. (1986).

In terms of performances, the expression of *kyoshuku* is certainly the most marked: it is done with a strong bowing of the head (and torso) that is a typical movement in relation to polite behavior, and is found (albeit in a much-reduced fashion) in *teinei* and *seii* – and in *heijo* if one looks at the rapid eye downward movement that is observed in this expression. This strong bow goes with a frowning face that is metaphoric of the suffering supposedly felt by the speaker, and that serves the negative politeness display used to alleviate the imposition done while using this expression. Acoustically, *kyoshuku* presents a flat intonation contour, that goes with a tense and harsh voice quality (see Shochi 2008) that also participates in the expression of suffering, and that is misperceived or misclassified in cross-cultural tests based on the audio modality only (Shochi 2008; but see Rilliard et al. 2009 and Shochi et al. 2020 for audio-visual disambiguation).

*Zonzai* is marked by an upward head movement, while looking at the addressee; its acoustic performance is relatively aroused, with intensity being slightly stronger than the others, while  $F_0$  is kept low. This inverse correlation of the two measures calls for a deliberate action so as to perform a rather lower voice, that would fit with predictions of the frequency code (Ohala 1983, 1994).

Inversely, *teinei* and *seii* show a large pitch span, with a high initial pitch that goes down regularly along the sentence; this  $F_0$  peak is accompanied by lower intensity. Here again, there is an inverse correlation between effort and  $F_0$ , with a high  $F_0$  target that matches the frequency code prediction, while the large pitch span, a mark of interest, is coherent with the polite behavior, and the predictions of the effort code (Gussenhoven 2004).

This work investigated the perception of five culturally encoded expressions of im/politeness in Japanese at various educational stages of children, using two experimental approaches. First of all, despite the differences in the paradigms, results were coherent, and indicate similar tendencies: the complexity of a cultural im/politeness system which does not fit a courtesy/discourtesy dimension, and requires a fine-grained social hierarchy to function. This knowledge is acquired progressively by children until reaching a highly coherent pattern among adults, as shown by Swerts (2011); Armstrong et al. (2018); Hübscher et al. (2017); Hübscher, Wagner, and Prieto (2020), and Long (2010), among others. An important multimodal effect was observed for an adequate description of im/polite expressions. The synergy of both audio and visual information helps to characterize the politeness meanings expressed by the speaker. For instance, Japanese polite behaviors (seii, teinei and kyoshuku) are characterized by a bowing of the head. In addition to this visual information, teinei and seii show similar acoustic characteristics (larger pitch span and decreased intensity, as compared to heijo). Due to this audio-visual similarity, these two expressions tend to be mixed up by children.

*Kyoshuku* was performed with a strong bowing of the head and torso, and with a frowned face that shows, together with a tense and harsh voice, the suffering of the speaker, as a strategy of negative politeness. On the contrary, *zonzai* showed an upward head movement while staring at the addressee, produced with a louder voice than the polite expressions.

In summary, we found that Japanese children learn to use and recognize im/politeness expressions in a socially adequate fashion between 6 to 10 years old, which demonstrates the growing cultural coherence gained with age. These expressions of im/politeness are organized in a complex multidimensional social scheme.

#### Funding

This research was partly supported by the Major Program of the National Social Science Fund of China (13&ZD189).

#### Acknowledgments

We are deeply grateful to children of Kurokami and Izumigaoka elementary schools as well as their teachers and the directors of these schools, to allow us to present our experiments. We also thank the adult listeners who participated in this experiment, and, especially, T. Ohmae. Much of the development of this research took place while the first author was a visiting researcher at Kyoto university thanks to the collaboration with Professor T. Sadanobu.

# References

- Armstrong, Meghan, Núria Esteve Gibert, Iris Hübscher, Alfonso Igualada, and Pilar Prieto. 2018. "Developmental and Cognitive Aspects of Children's Disbelief Comprehension through Intonation and Facial Gesture." *First Language* 38 (6): 596–616. https://doi.org/10.1177/0142723718789278
- Baltrušaitis, Tadas, Marwa Mahmoud, and Peter Robinson. 2015. "Cross-dataset Learning and Person-Specific Normalisation for Automatic Action Unit Detection." In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (*FG*), 6:1–6. IEEE. https://doi.org/10.1109/FG.2015.7284869
- Baltrušaitis, Tadas, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. "OpenFace 2.0: Facial Behavior Analysis Toolkit." In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59–66. Xi'an: IEEE. https://doi.org/10.1109/FG.2018.00019
- Benedict, Ruth. 1947. *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. London: Secker & Warburg.
- Boersma, Paul, and David Weenink. 2020. *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.16* (version 6.1.16). http://www.praat.org/
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Studies in Interactional Sociolinguistics 4. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085
- Burdelski, Matthew. 2010. "Socializing Politeness Routines: Action, Other-Orientation, and Embodiment in a Japanese Preschool." *Journal of Pragmatics* 42 (6): 1606–21. https://doi.org/10.1016/j.pragma.2009.11.007
- Calvetti, Paolo. 2020. "Strategies of Impoliteness in Japanese Spontaneous Talks." In *European Approaches to Japanese Language and Linguistics*, ed. by Giuseppe Pappalardo, and Patrick Heinrich, 13: 99–114. Ca' Foscari Japanese Studies. Venice: Fondazione Università Ca' Foscari. https://doi.org/10.30687/978-88-6969-428-8/005
- Culpeper, Jonathan. 1996. "Towards an Anatomy of Impoliteness." *Journal of Pragmatics* 25 (3): 349–67. https://doi.org/10.1016/0378-2166(95)00014-3
- Culpeper, Jonathan. 2010. "Conventionalised Impoliteness Formulae." *Journal of Pragmatics* 42 (12): 3232–45. https://doi.org/10.1016/j.pragma.2010.05.007
- Culpeper, Jonathan. 2016. "Impoliteness Strategies." In *Interdisciplinary Studies in Pragmatics, Culture and Society*, ed. by Alessandro Capone and Jacob L. Mey, 4: 421–45. Perspectives in Pragmatics, Philosophy & Psychology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12616-6\_16
- Culpeper, Jonathan, Derek Bousfield, and Anne Wichmann. 2003. "Impoliteness Revisited: With Special Reference to Dynamic and Prosodic Aspects." *Journal of Pragmatics* 35 (10–11): 1545–79. https://doi.org/10.1016/S0378-2166(02)00118-2
- Culpeper, Jonathan, Michael Haugh, and Dániel Z. Kádár (eds). 2017. *The Palgrave Handbook* of Linguistic (Im)Politeness. London: Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-37508-7
- Ekman, Paul, Wallace V. Friesen, and Joseph C. Hager. 2002. *Facial Action Coding System: The Manual*. Salt Lake City, Utah: Research Nexus.

Erickson, Donna, Shigeto Kawahara, Albert Rilliard, Ryoko Hayashi, Toshiyuki Sadanobu, Yongwei Li, Hayato Daikuhara, João de Moraes, and Kerrie Obert. 2020. "Cross Cultural Differences in Arousal and Valence Perceptions of Voice Quality." In *Speech Prosody 2020*, 720–24. ISCA. https://doi.org/10.21437/SpeechProsody.2020-147

Fernández-Dols, José-Miguel, and James A. Russell (eds). 2017. *The Science of Facial Expression*. Oxford Series in Social Cognition and Social Neuroscience. New York: Oxford University Press. https://doi.org/10.1093/acprof:0s0/9780190613501.003.0024

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. 1st ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

原田大樹 [Harada, Hiroki]. 2011. "敬語指導の現状と課題:小学校国語科を中心に. [Current Status and Issues honorific study: focusing on elementary school Japanese language department]." 日本教科教育学会. https://doi.org/10.18993/jcrdajp.34.3\_21

Hill, Beverly, Sachiko Ide, Shoko Ikuta, Akiko Kawasaki, and Tsunao Ogino. 1986. "Universals of Linguistic Politeness: Quantitative Evidence from Japanese and American English." *Journal of Pragmatics* 10 (3): 347–71. https://doi.org/10.1016/0378-2166(86)90006-8

Hong, Min-p'yo. 1993. "丁寧表現に見られる日本語音声の韻律的特徴 [Prosodic Characteristics of Japanese Polite Speech (in Japanese)]." *The Bulletin of the Phonetic Society of Japan* 204: 35-40.

Hübscher, Iris, Núria Esteve-Gibert, Alfonso Igualada, and Pilar Prieto. 2017. "Intonation and Gesture as Bootstrapping Devices in Speaker Uncertainty." *First Language* 37 (1): 24–41. https://doi.org/10.1177/0142723716673953

Hübscher, Iris, Laura Wagner, and Pilar Prieto. 2020. "Three-Year-Olds Infer Polite Stance from Intonation and Facial Cues." *Journal of Politeness Research* 16 (1): 85–110. https://doi.org/10.1515/pr-2017-0047

Husson, François, Sébastien Lê, and Jérôme Pagès. 2017. *Exploratory Multivariate Analysis by Example Using R.* Second edition. Boca Raton: CRC Press. https://doi.org/10.1201/b21874

- Ide, Sachiko. 1989. "Formal Forms and Discernment: Two Neglected Aspects of Universals of Linguistic Politeness." *Multilingua: Journal of Cross-Cultural and Interlanguage Communication* 8 (2–3): 223–48. https://doi.org/10.1515/mult.1989.8.2-3.223
- Ide, Sachiko. 2002. "The Speaker's Viewpoint and Indexicality in a High Context Culture." In Bunka, Intaakushon, Gengo [Culture, Interaction, and Language], ed. by Kuniyoshi Kataoka, and Sachiko Ide, 3–20. Tokyo: Hitsuji Shobo.

Ide, Sachiko. 2006. Wakimae no goyoron. Tokyo: Taishukan shoten.

Idemaru, Kaori, Bodo Winter, and Lucien Brown. 2019. "Cross-Cultural Multimodal Politeness: The Phonetics of Japanese Deferential Speech in Comparison to Korean." *Intercultural Pragmatics* 16 (5): 517–55. https://doi.org/10.1515/ip-2019-0027

Idemaru, Kaori, Bodo Winter, Lucien Brown, and Grace Eunhae Oh. 2020. "Loudness Trumps Pitch in Politeness Judgments: Evidence from Korean Deferential Speech." *Language and Speech* 63 (1): 123–48. https://doi.org/10.1177/0023830918824344

Ito, Mika. 2002. "Japanese Politeness and Suprasegmentals: a Study Based on Natural Speech Materials." In *Proc. Speech Prosody 2002*, 415–18.

Izard, Carroll E. 2007. "Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm." *Perspectives on Psychological Science* 2 (3): 260–80. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Kawano, Toshiyuki. 1995. "プロソディ-と丁寧表現-東京・大阪・名古屋の方言差を考慮して [Relationship between Prosody and Politeness Expressions (in Japanese)]." *The Bulletin of the Phonetic Society of Japan* 208: 9–17.

© 2023. John Benjamins Publishing Company All rights reserved

- Kiyama, Sachiko, Katsuo Tamaoka, and Masato Takiura. 2012. "Applicability of Brown and Levinson's Politeness Theory to a Non-Western Culture: Evidence from Japanese Facework Behaviors." *SAGE Open* 2 (4): 215824401247011. https://doi.org/10.1177/2158244012470116
- Kumbasar, Ece, A. Kimball Romney, and William H. Batchelder. 1994. "Systematic Biases in Social Perception." *American Journal of Sociology* 100 (2): 477–505. https://doi.org/10.1086/230544
- Long, Christopher. 2010. "Apology in Japanese Gratitude Situations: The Negotiation of Interlocutor Role-Relations." *Journal of Pragmatics* 42 (4): 1060–75. https://doi.org/10.1016/j.pragma.2009.095
- Matsumoto, Yoshiko. 1988. "Reexamination of the Universality of Face: Politeness Phenomena in Japanese." *Journal of Pragmatics* 12 (4): 403–26. https://doi.org/10.1016/0378-2166(88)90003-3
- Max, Ludo, and Patrick Onghena. 1999. "Some Issues in the Statistical Analysis of Completely Randomized and Repeated Measures Designs for Speech, Language, and Hearing Research." *Journal of Speech, Language, and Hearing Research* 42 (2): 261–70. https://doi.org/10.1044/jslhr.4202.261
- Mixdorff, Hansjörg, Angelika Hönemann, Albert Rilliard, Tan Lee, and Matthew K.H. Ma. 2017. "Audio-Visual Expressions of Attitude: How Many Different Attitudes Can Perceivers Decode?" *Speech Communication* 95 (December): 114–26. https://doi.org/10.1016/j.specom.2017.08.009
- Moore, Carmella C., A. Kimball Romney, Ti-Lien Hsia, and Craig D. Rusch. 1999. "The Universality of the Semantic Structure of Emotion Terms: Methods for the Study of Interand Intra-Cultural Variability." *American Anthropologist* 101 (3): 529–46. https://doi.org/10.1525/aa.1999.101.3.529
- Nadeu, Marianna, and Pilar Prieto. 2011. "Pitch Range, Gestural Information, and Perceived Politeness in Catalan." *Journal of Pragmatics* 43 (3): 841–54. https://doi.org/10.1016/j.pragma.2010.09.015
- Ofuka, Etsuko, J. Denis McKeown, Mitch G. Waterman, and Peter J. Roach. 2000. "Prosodic Cues for Rated Politeness in Japanese Speech." *Speech Communication* 32 (3): 199–217. https://doi.org/10.1016/S0167-6393(00)00009-1
- Ohala, John J. 1983. "Cross-Language Use of Pitch: An Ethological View." *Phonetica* 40 (1): 1–18. https://doi.org/10.1159/000261678
- Ohala, John J. 1994. "The Frequency Code Underlies the Sound-Symbolic Use of Voice Pitch." In *Sound Symbolism*, ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala, 325–47. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511751806.022
- Ohara, Yumiko. 2001. "Finding One's Voice in Japanese: A Study of the Pitch Levels of L2 Users." In *Multilingualism, Second Language Learning, and Gender*, ed. by Aneta Pavlenko, Adrian Blackledge, Ingrid Piller, and Marya Teutsch-Dwyer. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110889406.231
- Rao, C. R., and S. Suryawanshi. 1996. "Statistical Analysis of Shape of Objects Based on Landmark Data." *Proceedings of the National Academy of Sciences* 93 (22): 12132–36. https://doi.org/10.1073/pnas.93.22.12132
- Rilliard, Albert, Donna Erickson, João Antônio de Moraes, and Takaaki Shochi. 2014. "Cross-Cultural Perception of Some Japanese Politeness and Impoliteness Expressions." In *Linguistic Approaches to Emotions in Context*, ed. by Fabienne Baider, and Georgeta Cislaru, 251–76. Pragmatics & Beyond New Series 241. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.241.15ril

© 2023. John Benjamins Publishing Company All rights reserved Rilliard, Albert, Donna Erickson, João Antônio de Moraes, and Takaaki Shochi. 2017. "Perception of Expressive Prosodic Speech Acts Performed in USA English by L1 and L2 Speakers." *Journal of Speech Sciences* 6 (1): 27–45. https://doi.org/10.20396/joss.v6i1.14981

Rilliard, Albert, Takaaki Shochi, Jean-Claude Martin, Donna Erickson, and Véronique Aubergé. 2009. "Multimodal Indices to Japanese and French Prosodically Expressed Social Affects." *Language and Speech* 52 (2–3): 223–43. https://doi.org/10.1177/0023830909103171

Romney, A. K., J. P. Boyd, C. C. Moore, W. H. Batchelder, and T. J. Brazill. 1996. "Culture as Shared Cognitive Representations." *Proceedings of the National Academy of Sciences* 93 (10): 4699–4705. https://doi.org/10.1073/pnas.93.10.4699

Romney, A.K., C.C. Moore, W.H. Batchelder, and T.-L. Hsia. 2000. "Statistical Methods for Characterizing Similarities and Differences between Semantic Structures." *Proceedings of the National Academy of Sciences* 97 (1): 518–23. https://doi.org/10.1073/pnas.97.1.518

Romney, A.K., C.C. Moore, and C.D. Rusch. 1997. "Cultural Universals: Measuring the Semantic Structure of Emotion Terms in English and Japanese." *Proceedings of the National Academy of Sciences* 94 (10): 5489–94. https://doi.org/10.1073/pnas.94.10.5489

Roseberry-McKibbin, Celeste, and Alejandro Brice. 1999. "The Perception of Vocal Cues of Emotion by Spanish-Speaking Limited English Proficient Children." *Journal of Children's Communication Development* 20 (2): 19–24. https://doi.org/10.1177/152574019902000203

Sadanobu, Toshiyuki. 2004. "A Natural History of Japanese Pressed Voice." *Journal of the Phonetic Society of Japan* 8 (1): 29–44. https://doi.org/10.24467/onseikenkyu.8.1\_29

Scherer, Klaus R. 2001. "Appraisal Considered as a Process of Multi-Level Sequential Checking." In Appraisal Processes in Emotion: Theory, Methods, Research, ed. by Klaus R. Scherer, Angela Schorr, and Tom Johnstone, 92–120. Oxford: Oxford University Press.

Scherer, Klaus R. 2009. "The Dynamic Architecture of Emotion: Evidence for the Component Process Model." *Cognition & Emotion* 23 (7): 1307–51. https://doi.org/10.1080/02699930902928969

Scherer, Klaus R. 2013. "Measuring the Meaning of Emotion Words: A Domain-Specific Componential Approach." In *Components of Emotional Meaning*, ed. by Johnny R. J. Fontaine, Klaus R. Scherer, and Cristina Soriano, 7–30. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:0s0/9780199592746.003.0002

Sherr-Ziarko, Ethan. 2019. "Prosodic Properties of Formality in Conversational Japanese." Journal of the International Phonetic Association 49 (3): 331–52. https://doi.org/10.1017/S0025100318000117

Shochi, Takaaki. 2008. "Prosodie des affects socioculturels en japonais, et anglais : à la recherche des vrais et faux-amis pour le parcours de l'apprenant." PhD Thesis, Université Stendhal, Grenoble III.

- Shochi, Takaaki, Véronique Aubergé, and Albert Rilliard. 2007. "Cross-Listening of Japanese, English and French Social Affect: About Universals, False Friends and Unknown Attitudes." In Proceedings of the 16th International Congress of Phonetic Sciences, 2097–2100. Saarbrucken, Germany.
- Shochi, Takaaki, Donna Erickson, Albert Rilliard, Véronique Aubergé, and Jean-Claude Martin. 2008. "Recognition of Japanese Attitudes in Audio-Visual Speech." In 4th International Conference on Speech Prosody, 689–92. Campinas, Brazil.

- Shochi, Takaaki, Donna Erickson, Albert Rilliard, Kaoru Sekiyama, and Véronique Aubergé. 2009. "Japanese Children's Perception for Audio-Visual Prosody of Social Politeness." In National Conference on Man-Machine Speech Communication: International Symposium on Asian Speech Resources, 6p. Xinjiang Province, China.
- Shochi, Takaaki, Marine Guerry, Albert Rilliard, Donna Erickson, and Jean-Luc Rouas. 2020. "The Combined Perception of Socio-affective Prosody: Cultural Differences in Pattern Matching." *Journal of the Phonetic Society of Japan* 24 (December): 84–96. https://doi.org/10.24467/onseikenkyu.24.0\_84
- Shochi, Takaaki, Takeki Kamiyama, Albert Rilliard, and Véronique Aubergé. 2014.
  "Apprentissage des expressions prosodiques et gestuelles de politesse par des francophones." In *Histoires d'amour : quelques modalités de relation à l'autre au Japon*, ed. by Makiko Andro-Ueda, and Jean-Michel Butel, 417–424. Japon Pluriels 9. Paris: Philippe Picquier.
- Shochi, Takaaki, Albert Rilliard, Véronique Aubergé, and Donna Erickson. 2009. "Intercultural Perception of English, French and Japanese Social Affective Prosody." In *The Role of Prosody in Affective Speech*, ed. by Sylvie Hancil, 31–60. Linguistic Insights 97. Bern, Switzerland: Peter Lang AG.
- Spencer-Oatey, Helen. 1996. "Reconsidering Power and Distance." *Journal of Pragmatics* 26 (1): 1–24. https://doi.org/10.1016/0378-2166(95)00047-X
- Swerts, Marc. 2011. "Correlates of Social Awareness in the Visual Prosody of Growing Children." *Laboratory Phonology* 2 (2): 381–402. https://doi.org/10.1515/labphon.2011.014
- Swerts, Marc, and Emiel Krahmer. 2005. "Audiovisual Prosody and Feeling of Knowing." *Journal of Memory and Language* 53 (1): 81–94. https://doi.org/10.1016/j.jml.2005.02.003
- Wierzbicka, Anna. 1985. "Different Cultures, Different Languages, Different Speech Acts." Journal of Pragmatics 9 (2-3): 145–78. https://doi.org/10.1016/0378-2166(85)90023-2
- Wierzbicka, Anna. 1996. "Japanese Cultural Scripts: Cultural Psychology and 'Cultural Grammar'." *Ethos* 24 (3): 527–55. https://doi.org/10.1525/eth.1996.24.3.02a00060
- Zinck, Alexandra, and Albert Newen. 2008. "Classifying Emotion: A Developmental Account." *Synthese* 161 (1): 1–25. https://doi.org/10.1007/s11229-006-9149-2