

Grammar, Pronunciation, or Something Else? Native Japanese Speakers' Judgments of "Native-Like" Speech

Mayumi Ajioka

University of California, Los Angeles

This paper explores speech factors that influence native Japanese speakers' perceptions of "native-like" speech. The conventional criterion of "native-like" proficiency has focused on grammar or pronunciation, which researchers recognize as important. This paper challenges this top-down discussion of "native-likeness" and examines the online (while listening) and offline (after listening) perceptions of 108 native Japanese speakers who are not academic researchers in a multi-dimensional way, in order to investigate (1) what factor(s) contribute to perceptions of "native-like" speech? and (2) For linguistically lay people, what factors determine "native-like" speech?

The methods of analysis used were factor analysis and correlations. My analysis of online perceptions of "native-likeness" is consistent with prior research that highlights grammar and pronunciation as the most important and noticeable features of non-native speakers' speech. However, my analysis of offline perceptions reveals the significance of interaction-related factors, suggesting that grammar and pronunciation are less influential on native speakers' holistic judgment of "native-like" speech. From these results, I propose two types of unnaturalness: overt and covert, the latter of which is illustrated to have a profound effect on native speakers' overall impressions of non-native speakers' speech. In conclusion, this paper highlights a possible disagreement between academic and lay perspectives with implications for teaching that places more emphasis on interaction than on accuracy for L2 learners.

Introduction

This paper aims to discuss the influential factors of native Japanese speakers' judgments of "native-like" speech. By analyzing perceptions in multi-faceted analyses, the paper explores what element(s) make(s) non-native speaker's speech sound "native-like" or "non-native-like" to native Japanese speakers. Different research fields have recognized different elements as key to native speakers' perceptions of L2 speakers' "native-like" speech. Second Language Acquisition (SLA) experts, like Johnson and Newport (1989) and Flege (1999), have considered grammar or pronunciation to be the most important variables that impact non-native speakers' "native-like" proficiency. Alternatively, functional linguists, like Fillmore (1979) and Pawley and Syder (1983a, 1983b, 2000), have taken a strong interest in speakers' pauses and hesitations. These different views prompt two questions: (1) what *is* the most important factor? And, (2) do lay people agree with the key elements of "native-like" speech that scholars put forward?

This study investigated 108 native Japanese speakers' perceptions of non-native speakers' speech. The method was specifically designed to examine two perceptions: *online* and *offline*. *Online* perception represents native speakers' immediate perception *while* they listen to a non-native speaker's utterances, whereas *offline* perception indicates their subsequent perception *after* they have finished listening to the data. By analyzing these two perceptions, the study offers a nuanced understanding of "native-like" speech.

The results show that grammar and pronunciation has less influence on native speakers' judgments of "native-like" speech as opposed to the conventional emphasis on these two categories. Instead, smooth interaction and mutual understanding seems to have more importance on native speakers' perceptions. Moreover, I demonstrate how two different types of unnaturalness affected *online* and *offline* perceptions, which emphasizes the significance that interaction-related factors have on native speakers' judgments of "native-like" speech.

Two Fundamental Questions from Past Studies

Research in the fields of SLA and functional linguistics has put forth separate approaches and discussions regarding the issue of "native-like" proficiency and fluency. These discrepancies inevitably cast the first fundamental question on this issue: *What factor(s) contribute to perceptions of "native-like" speech?*

In the area of SLA, the topic of "native-like" proficiency had been chiefly investigated in relation to the Critical Period Hypothesis (CPH). This hypothesis posits that successful language acquisition occurs between birth and the onset of puberty (the Critical Period), after which time, language acquisition becomes more difficult and ultimately less successful (Lenneberg, 1967). This hypothesis is only concerned with first language (L1) acquisition, but soon after its publication, a host of SLA researchers began to apply this hypothesis to second language (L2) acquisition and examine the relationship between age and success in acquiring "native-like" proficiency of the target language.

The methods used most frequently to measure learners' "native-like" proficiency were pronunciation tests (Bongaerts, 1999; Flege, Munro, & MacKay, 1995) and sentence grammaticality judgment tests (DeKeyser, 2000; Johnson & Newport, 1989). In the pronunciation tests, L2 learners were asked to pronounce a given L2 word or sentence in isolation while examiners (native speakers) judged their pronunciations in terms of closeness to a native speaker's pronunciation. In the grammaticality judgment tests, L2 learners were asked to judge the grammaticality of several sentences, some of which included a grammatical deviation. Thus, pronunciation and grammar have traditionally been considered to be the most important components in measuring non-native learners' "native-like" proficiency in the field of SLA.

In functional linguistics, the issue of “native-like” speech or fluency has long been discussed in terms of a more natural human question, “*What features distinguish native and non-native speakers’ speech?*” or “*What features determine “nativeness?”*” The most remarkable studies were conducted by Andrew Pawley and Frances H. Syder (1983a, 1983b, 2000). They proposed two characteristics as key elements that mark “nativeness” in English: (1) “one clause at a time *facility*,” meaning, a fluent speaker regularly encodes one clause and avoids the need for mid-clause pauses or hesitations; and (2) “memorized sentence” and “lexicalized sentence stem,” which implied that native speakers somehow select a type of fixed expression out of various grammatical paraphrases. In sum, functional linguists have focused on the use of fixed phrases and the position of pauses or hesitations as significant factors of “native-like” speech, not grammar or pronunciation.

In both research areas, key components of “native-like” proficiency or speech were determined by researchers’ perspectives, including: grammar and pronunciation in SLA and fixed phrases and pauses/hesitation in functional linguistics. However, if non-native learners go into their target culture, their conversation partners will likely not be language researchers, but those they interact with on a daily basis in a wide range of capacities. This raises a second, perhaps more fundamental question: *For non-researchers, what factors determine “native-like” speech?*

This study addresses both fundamental questions with two major contributions to the study of “nativeness” in second language acquisition. Firstly, I focus on judgments of “nativeness” beyond the “researchers” perspective to include participants with no formal knowledge of language assessment. Additionally, I focus on judgments of Japanese speech, in the first study of its kind. Specifically, the study seeks to answer the following question: “*What factors of speech do native speakers use to judge the “native-likeness” of the learner’s speech?*” To explore the issue further, I posed three research questions:

RQ1: Do the native speakers’ perceptions agree with grammaticality?

RQ2: Do their perceptions focus on pronunciation?

RQ3: Are there any other important factors that must be considered?

Study Methods

The study asked 108 native Japanese speakers living in or around Tokyo to listen to non-native speakers’ interview data and make a judgment on their “native-likeness”. These 108 participants were neither linguistic specialists nor SLA researchers. Of these 108 participants, 106 were undergraduate students and two were housewives in their thirties. Their past experiences of foreign languages and cultures and their amount of exposure to learner Japanese vary from person to person, though many remain monolingual in Japanese. Some of them have an experience of living abroad, mostly with their families accompanying their fathers on extended business stays for periods up to four or five years.

The procedure was as follows. I distributed a study packet to each of the participants, which included the following five items: (1) a CD of four learners' interviews in Japanese; (2) transcripts (in Japanese) of the four interviews; (3) four questionnaire sheets on the four interviewees' speech; (4) a personal information sheet; and (5) a consent form.

The four interviews were all given to non-native Japanese speakers by native Japanese speakers. I carefully selected the four interviews out of the *UEMURA CORPUS* (Uemura, 1998) so that each interviewee might have different ranges of strengths and weaknesses. In order to make clear which element(s) most influenced native speakers' judgment of "native-likeness" of non-native speakers' speech, I selected the four interviews with varying degrees of proficiency with respect to different aspects of the language.

The following four non-native speakers' interviews were used for the study (see also Table 1; the names given here are pseudonyms): (1) Ami, a female Korean, who had fewer grammar errors and made longer expressions and narratives than the other interviewees but sometimes included Korean pronunciation of Japanese. Her interview responses stuck to the prompt and rarely developed into a conversation; (2) Kay, a female American, who had shorter answers than Ami, relied frequently on English expressions and vocabulary and despite occasional grammatical errors was able to actively engage the interviewer in a more conversational style of an interview; (3) Ben, a male American, who had a large academic Japanese vocabulary and academic exposure to the language through his graduate studies sometimes had false starts or repeated the same fillers while he was searching for a word or expression; and finally (4) Dan, a male Korean-American, who had a smaller vocabulary, more grammar errors and hesitations, and less clear articulation than the other three, but who had the most Japanese-like pronunciation and intonation with fewer repetitive fillers than Ben. I controlled the interviewees' proficiency levels largely ranging from high intermediate to low advanced, because learners above this level have almost no foreign accent left in their speech while those below this level do not have an ability to maintain a conversation.

Table 1.

Information on the Four Interviewees' Japanese Proficiency

	Language Background	Sex	Pronunciation	Grammar	Vocabulary	Others
Ami	Korean	F	-	+		long speech
Kay	English	F			-	actively engaging
Ben	English	M			+	false starts, many fillers
Dan	English/ Korean (home)	M	+	-	-	hesitations, unclear articulation

Note. (In the table, "- (minus)" indicates a learner's notable weakness, while "+ (plus)" indicates his or her notable strength.

I asked the participants to listen to each of the four interviewees and to perform the following three tasks: (1) *while* listening to each interview, highlight any part of the transcript that did not sound “native-like” or natural in any way (*online* perception); (2) *after* listening to each interview, fill out a questionnaire on each interviewee’s speech (*offline* perception); and (3) *after* listening to all four interviews, rank the four interviewees’ speech in a “native-like” or natural order (*offline* judgment).

The highlighted transcripts (*online* perception) were analyzed by tagging the transcripts and using a concordance tool, which made it easy to find which word or tag most frequently occurred or which word(s) or phrase(s) most frequently came before or after a particular word or tag. In this study, I tagged every highlighted part of the collected data with an open tag <unnatural> and a close tag </unnatural> so that I could easily see between the tags (i.e., the highlighted part) and how many participants highlighted the same part. After that, four Japanese linguists categorized all the highlighted parts according to types of unnaturalness, such as grammar and pronunciation (for seven other categories and their descriptions, see Ajioka 2008).

On the other hand, the collected questionnaires (*offline* perception) were analyzed by means of factor analysis. This analysis is often used in psychology and the social sciences, wherein a researcher asks a number of participants to fill out a questionnaire and a statistical tool is used to find unobserved potential factors underlying the responses to the question items. My questionnaire contained 45 question items¹, all in Japanese, and answers were based on a 5-point scale ranging from 1 (poor) to 5 (good).

Findings

In this section, I summarize the findings of both questionnaires (*offline* perception) and highlighted transcripts (*online* perception).

Questionnaires Analyzed by Factor Analysis²

First, one can examine the factor-analysis findings of the questionnaire that examined the *offline* perception of Japanese participants. Using a statistical program, I extracted four factors – *Fluency*, *Comprehension*, *Lexical/Grammatical*, and *Phonological/Phonetic*³ – from the question items (observed variables). The obtained factors are shown below in Table 2.

Factors 1 and 2 are both related to interaction with others; in other words, the question items grouped in these factors are only possible when the speaker has a conversation partner. On the other hand, Factors 3 and 4 are both associated with linguistic elements; the items included in these two factors all describe the speech itself whether or not the speaker has a conversation partner. With respect to interactive factors, comparing Factor 1 with Factor 2, the items of Factor 1 – *Fluency* – include pauses, fillers, hesitations, and silence, which are related to the learner’s fluency and speech production, i.e., what was expressed from the learner.

In contrast, the components of Factor 2 – *Comprehension* – are more concerned with how much the non-native interviewee comprehends the native interviewer’s question and answers appropriately, though they are still related to interaction with the interviewer. Likewise, looking at Factors 3 and 4 more closely, the components of Factor 3 – *Lexical/Grammatical* – are all concerned with the learner’s vocabulary and grammar while those of Factor 4 – *Phonological/Phonetic* – are related to the learner’s pronunciation and articulation.

Table 2.

Pattern Matrix of Factor Analysis

	Factor 1	Factor 2	Factor 3	Factor 4
	Fluency	Compre- hension	Lexical / Grammati- cal	Phonologi- cal / Pho- netic
Natural Frequency of Fillers	.951	-.070	-.236	.017
Natural Position of Pauses	.906	-.123	.097	-.036
Natural Frequency of Pauses	.874	-.069	.022	-.044
Natural Position of Fillers	.857	.010	-.154	-.037
Speaking without False Starts	.760	.025	-.026	.050
Speaking without Much Effort	.655	.115	.110	-.064
Speaking without Long Silence	.635	.060	.142	.012
Speaking without Redundant Use of Words	.611	.112	-.016	.044
Smooth Flow of Speech	.443	.321	.080	.036
Natural Speed of Speech	.394	.147	.150	.082
Speaking without Circumlocution	.386	.096	.134	.125
Comprehending the Questions Correctly	-.199	.990	-.006	.003
Answering the Questions Correctly	.054	.918	-.164	.008
Comprehending the Questions Quickly	-.068	.900	.036	-.131
Making His/Her Point Clear	.116	.774	-.079	-.021

Responding to Any Topic Flexibly	.105	.766	-.116	.068
Making Coherent/Consistent Statements	.088	.725	-.132	.091
Responding Earnestly	-.017	.467	.211	.110
Answering Even Difficult Questions without Giving Up	.128	.420	.281	-.075
Responding without Too Many Comprehension Checks	.082	.394	.284	-.040
Speaking without L1 Word Mixing	-.153	-.172	.984	-.033
Speaking without Direct Translation from L1	-.011	-.187	.878	-.002
Having a Knowledge of Difficult Japanese Words	-.035	-.008	.828	-.093
Having a Large Size of Japanese Vocabulary	-.011	.136	.721	-.035
Using Keigo (Honorifics) Naturally	.011	.027	.609	.123
Using the Expression Appropriate for the Situation	-.032	.259	.548	.025
Using Natural Sentence-Final Particles	.158	-.002	.440	.260
Using Formulaic Expressions Properly	.298	.007	.435	-.041
Natural Word Accent	.042	-.068	-.147	.927
Natural Sentence Intonation	.044	-.052	-.036	.884
Natural Pronunciation	.052	-.068	-.038	.827
Natural Sokuon (the Pronunciation of Double Consonant)	-.050	.008	.084	.707
Natural Dakuon (the Pronunciation of Muddy Sound)	-.149	.142	.037	.643
Clear Articulation	.050	.111	.192	.532
Extraction Method:	Principal Axis Factoring			
Rotation Method:	Promax with Kaiser Normalization			

The values of factor contribution of these four factors are: 12.180 (*Fluency*), 11.476 (*Comprehension*), 9.224 (*Lexical/Grammatical*), and 9.842 (*Phonological/Phonetic*). When each factor is correlated to each other, as are the factors of this study (Table 3), we should be careful in taking these values into consideration, because these values are obtained by ignoring the factor correlations. However, it may be safe to say that Factors 1 and 2 – *Fluency* and *Comprehension*, respectively – have higher values than Factors 3 and 4 – *Lexical/Grammatical* and *Phonological/Phonetic*, respectively. In addition, Table 3 shows that Factors 1 and 2 have the highest correlations of all ($r = .713$).

Table 3.
Factor Correlation Matrix.

Factor	Fluency	Comprehension	Lexical / Grammatical	Phonological / Phonetic
Fluency	1			
Comprehension	.713	1		
Lexical / Grammatical	.555	.548	1	
Phonological / Phonetic	.669	.590	.525	1

The results of the factor analysis are summarized as follows:

1. Four factors are extracted of Japanese native speakers' *offline* judgments of non-native speakers' speech: *Fluency*, *Comprehension*, *Lexical/Grammatical*, and *Phonological/Phonetic*.

2. The *Fluency Factor* (regarding speech production during conversation) had the biggest influence on native Japanese speakers' perception or judgment of "native-like" speech, followed by the *Comprehension Factor* (regarding comprehension and attitude during conversation).

3. The interaction-related factors (*Fluency* and *Comprehension*) seemed to have more influence than the linguistic factors (*Lexical/ Grammatical* and *Phonological/Phonetic*).

Highlighted Transcripts and Rankings Analyzed by Correlation

In this section, I summarize the important findings of analyzing 105 highlighted transcripts⁴ (*online* perception) and the rankings that native speakers gave to each of the non-native interviewees (*offline* judgment). Table 4, below, provides

a summary of numbers and percentages of each interviewee’s highlighted parts and their average ranking scores. Only the most remarkable categories were excerpted. For a full table that includes all categories, see Ajioka (2008).

Table 4.
*The Numbers and Percentages of Highlighted Parts and Average Ranking Scores*⁵.

Interviewee	Ranking	Grammar	Pronunciation	Pragmatics	Disfluency	Total
Ami	1.58	308 13%	836 35%	77 3%	27 1%	2419 100%
Kay	2.10	782 25%	550 18%	497 16%	133 4%	3089 100%
Ben	2.63	711 23%	644 21%	242 8%	113 4%	3094 100%
Dan	3.68	1015 28%	836 23%	152 4%	535 15%	3669 100%

According to Ajioka (2008), the two most highlighted categories for all non-native interviewees, were *Grammar* and *Pronunciation*. This suggests that deviations of grammar and pronunciation were the easiest for native Japanese speakers to objectively identify. For Ami and Ben, the other categories were not significant enough to make a difference, comprising only 8% or less of noticeable “unnaturalness” tags. For Kay and Dan, however, the other categories were significant, as participants highlighted *Pragmatics* as the third most noticeable error for Kay, and *Disfluency* for Dan. For a detailed discussion of the relevance of these categories, see Ajioka (2008).

Examining the highlighted errors alone would lead one to conclude that previous research from SLA and Functional Linguistics does hold some ground. However, examining participants’ subjective ranking scores reveals some complexities with regard to how native speakers of Japanese perceive non-native speaker “naturalness,” at least in this local Tokyo context. Upon closer examination, we can see that the correlation between the average *Ranking* scores and the numbers of highlighted parts of each category (see Table 5) reveal some interesting findings. Based on the logic of SLA and Functional Linguistics research, one would expect that a higher frequency of “errors” highlighted (unnaturally sounding parts) would lead to a lower ranking.

Table 5.
*The Correlations Between Ranking and Each Category.*⁶

		Rank	Gram- mar	P r o - n u n - c i a t i o n	Disflu- ency	P r a g - m a t i c s	Total
Rank	Correlation	1					
	Sig. (2-tailed)	.					
	N	420					
Gram- mar	Correlation	.409	1				
	Sig. (2-tailed)	.000	.				
	N	420	420				
Pronun- c i a t i o n	Correlation	.107	.165	1			
	Sig. (2-tailed)	.028	.001	.			
	N	420	420	420			
Disflu- ency	Correlation	.438	.595	.232	1		
	Sig. (2-tailed)	.000	.000	.000			
	N	420	420	420	420		
P r a g - m a t i c s	Correlation	.064	.440	.202	.312	1	
	Sig. (2-tailed)	.191	.000	.000	.000	.	
	N	420	420	420	420	420	
Total	Correlation	.260	.633	.696	.553	.576	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.
	N	420	420	420	420	420	420

Table 5, above, shows the correlations between each noteworthy category and *Ranking*. Each cell in Table 5 holds three values: correlation coefficient on the top, *p*-value in the middle, and the number of data on the bottom. The higher correlation coefficient yields an expected higher correlation, either positive or negative. In short, what matters here is the absolute value. The *p*-value indicates the statistical significance; the lower this value is, the more significant the correlation. I set the significant level as .05, following the convention and default settings of statistic program. Therefore, all the values of correlation coefficient are statistically significant, except the one between *Rank* and *Pragmatics* ($p = .191$).

Let us look at each correlation coefficient⁷ in Table 5. I describe the noteworthy results in the following four sections: (1) between *Rank* and the categories (the categories include *Grammar*, *Pronunciation*, *Disfluency*, and *Pragmatics*); (2) between each category; (3) between *Total* and the categories; and (4) between *Rank* and *Total*.

1. The Correlations Between *Rank* and the Categories:

The correlations between *Rank* and *Grammar* ($r = .409$) and between *Rank* and *Disfluency* ($r = .438$) are both moderate and the latter is a little higher than the former. The other correlation coefficients, *Rank* and *Pronunciation* ($r = .107$) and *Rank* and *Pragmatics* ($r = .064$; $p = .191$), are ignorable here; *Rank* and *Pronunciation* have only a very low correlation and *Rank* and *Pragmatics* have a very low and no significant correlation.

2. The Correlations Between Each Category:

The correlation between *Grammar* and *Disfluency* ($r = .595$) is moderate to relatively high and the correlation between *Grammar* and *Pragmatics* ($r = .440$) is moderate. This suggests that grammar errors are likely to go hand-in-hand with pragmatic unnaturalness and disfluency in non-native speakers' speech. On the other hand, *Disfluency* and *Pragmatics* have only a low correlation ($r = .312$), which indicates that the lack of grammar skills may be the main source of disfluency and pragmatic unnaturalness. However, it is also important to add that *Disfluency* has a higher correlation with *Rank* than *Grammar* with *Rank*; that is, even though grammar errors may trigger disfluency, the holistic judgment in native-like speech is slightly more related to fluency than to grammar itself.

Another interesting result is the correlations around *Pronunciation*. That is, no notable correlations were obtained from *Pronunciation*, which indicates that the naturalness (or unnaturalness) of pronunciation is entirely separated from other kinds of proficiency, with regard to the judgment of native-like speech: *Pronunciation* and *Grammar* are only very weakly correlated ($r = .165$); *Pronunciation* and *Disfluency* ($r = .232$) and *Pronunciation* and *Pragmatics* ($r = .202$) are weakly correlated.

3. The Correlations Between *Total* and the Categories:

All the correlations between *Total* and the Categories are in the range of moderate correlation: *Total* and *Grammar* ($r = .633$); *Total* and *Pronunciation* ($r = .696$); *Total* and *Disfluency* ($r = .553$); and *Total* and *Pragmatics* ($r = .576$). Also, two of them (*Total* and *Grammar* and *Total* and *Pronunciation*) can be described as at the level of moderate to relatively high correlation. This is natural if we consider the high number of highlightings on grammar errors and unnatural pronunciations, as is shown in Table 4. Therefore, the results of this section show that each of the

four categories contribute to the total number of highlightings, but the categories of *Grammar* and *Pronunciation* play the biggest part; i.e., these two categories of highlightings account most for the whole highlightings.

4. The Correlations Between *Total* and *Rank*:

The final and most interesting correlation is between *Rank* and *Total* ($r = .260$). There is only a low, and thus negligible, correlation between *Rank* and *Total*; that is, the participants' holistic evaluations of the four non-native speakers' Japanese speech, which are represented by rankings, do not correlate to their analytic evaluations of unnaturalness. This low correlation is presumably due to the high number of highlightings on unnatural pronunciation. In other words, unnatural pronunciation is easily noticeable to native speakers but does not affect their holistic judgment of native-like speech. This is also supported by the very weak correlation between *Rank* and *Pronunciation*. Furthermore, if we consider the moderate (not high) correlations between *Rank* and *Grammar* and between *Rank* and *Disfluency*, it can be assumed that the *online* perceptions of Japanese native speakers regarding non-native speaker's unnatural speech are not so highly related to their *offline* judgment of native-like speech.

In sum, the following results were found from the highlighted transcripts and the ranking scores that each of the non-native interviewees obtained.

1. *Pronunciation* have no immediate relation to the judgment of "native-like" speech.

2. *Grammar* have a moderate to relatively high correlation with *Disfluency* and a moderate correlation with *Pragmatics*.

3. *Disfluency* and *Grammar* have moderate correlations with *Rank*. (*Disfluency* has a slightly higher correlation than *Grammar*.)

4. *Total* has almost no notable relation to *Rank*

From the results through 2 – 4, if grammar errors are excessive and combine with other kinds of errors, it is more likely to affect native speakers' judgments on "native-like" speech. It seems basically, however, that native speakers' analytic evaluations do not always correspond to their holistic judgments.

Discussion:

Overt Unnaturalness vs. Covert Unnaturalness

In this section, I summarize all the results of this study and discuss what those results mean for native speakers' perception and judgment of non-native speakers' speech. The factor analysis of the questionnaires showed that the interaction-related factors—e.g., pauses, fillers, hesitations, etc. (the *Fluency Factor*) and comprehension, attitude, etc. (the *Comprehension Factor*)—have more influence on the judgment of "native-like" speech than the linguistic factors—i.e., the *Lexical/Grammatical Factor* and the *Phonological/Phonetic Factor*. This result suggests

that in the context of Japanese interaction in Tokyo, as long as both conversation partners establish communication and mutual understanding, grammar errors and pronunciation deviations do not count for as much as previous research has suggested.

The correlations of highlighted transcripts and rankings obtained partly similar but more interesting findings. The *Pronunciation* category of highlightings had almost no correlation with the *Ranking* scores that native speakers gave to each non-native interviewee. On the other hand, *Disfluency*, which has a close relation with the *Fluency Factor* in factor analysis, had a moderate correlation with *Ranking*. The *Grammar* category also had a moderate correlation with the *Ranking* scores, but if we look at the relatively high correlation of *Grammar* and *Disfluency*, we can see that these two features tend to cooccur in the same person. Finally, the *Total* number of highlightings and *Ranking* had only a weak correlation and *Ranking* had only moderate or lower correlations with each category. These results indicate that unnatural parts in non-native Japanese speech are easily noticeable to Japanese native speakers but they do not immediately influence the judgment as to whether the non-native speech is native-like.

Of course, these results from both analyses do not discount the import of learning grammar to get close to natural speech; a learner of Japanese with a large number of grammar errors is more likely to make errors in other aspects (which may indicate that this person is not fluent in the target language). Yet once a learner gains sufficient language experience and reaches a level of interactive proficiency when communicating with native speakers, grammar and pronunciation have less of an impact on perceptions of “native-like” Japanese speech. Disfluency, however, such as frequent or excessive pauses and fillers, seem to give native speakers an impression of “non-native-like” speech.

Based on the investigation of *offline* perceptions of native speakers, i.e., the holistic judgment or overall impression of each interviewee’s speech *after* listening to the interview, we can thus conclude that interaction-related factors are more influential on the judgment of “native-like” speech than grammar and pronunciation. On the other hand, native speakers’ *online* perception—i.e., the number and percentage of highlightings, which native speakers made *while* listening to the interview—shows another result: the two most highlighted categories were *Grammar* and *Pronunciation*, which conflicts with those of *offline* perceptions. Of the results of the investigation of *online* perception, grammar and pronunciation were the most remarkable and noticeable to native speakers, although they seem to have no immediate relations to native speakers’ overall judgments, i.e., the *Ranking* scores.

These results suggest that there are two types of unnaturalness in the perception of native Japanese speakers: *overt* and *covert*. *Overt unnaturalness* includes obvious grammatical errors, unnatural word choices, and pronunciation, which most native speakers can easily find it wrong or unnatural while they are listening to non-native speakers’ speech, as was demonstrated in the large number of highlightings classified as *Grammar* or *Pronunciation* in the collected transcripts.⁸

Overt unnaturalness, however, does not necessarily influence native speakers' final judgments or overall impressions of non-native speakers' speech. *Covert unnaturalness* involves disfluency, including hesitations and the position/frequency of pauses and fillers. Non-native speaker's sufficient comprehension and attitude to communicate actively are probably concerned with this type of unnaturalness, too. In sum, *covert unnaturalness* involves interaction-related factors. This categorization is not likely to be recognized as an objective error, but it may interfere with a smooth flow of communication and give an impression of "non-native-like" speech or an inexperienced non-native speaker to native speakers. While listening to non-native speech (*online* perception), *overt unnaturalness* may be easy to notice but does not remain in the native speakers' mind for long. More important and more influential to native speakers' overall impression toward non-native speech (*offline* perception) is *covert unnaturalness*, i.e., something that impedes smooth interaction.

The difference between these two types of unnaturalness in relation to *online/offline* perception results in two important implications, the first of which relates to academia. Conventionally, most language-related issues have been studied and discussed only from the perspective of academically trained linguists; grammaticality and pragmatic judgments have depended upon decisions by linguistically trained experts. However, language is neither owned nor dictated exclusively by academic specialists, but by all humanity, and, as shown in the results of this study, their views may disagree with the views of non-academic people (not to mention that the perceptions of non-academic people are much more complex than expected!). This suggests that the shift of focus to lay perception—i.e., *democratization of academics* of sorts—may well be instructive when dealing with "native-like" speech or fluency in everyday communication and interpersonal relationships.

The second implication is a pedagogical one. The overall results of this study showed that the goal for language learners should not be *perfect* accuracy of the target language, but rather smooth conversation for mutual understanding. *Overt unnaturalness* may be easy to pick up but if it does not impinge on the flow of the conversation, it does not count for much as even native speakers do not always adhere to prescriptive grammar. What hinders interaction and mutual understanding and evokes an impression of "non-native-like" or a lack of fluency is *covert unnaturalness*. This type of unnaturalness continues even after a conversation when the fundamental impression of a speech or speaker has been created. *Covert naturalness* is acquired only through actual interaction. Of course, I do not mean to imply that accuracy is unnecessary. Indeed, accuracy is essential until a learner reaches a certain level and interaction with a native speaker is possible to some degree. If a learner, however, is overly concerned with grammatical errors and pronunciation deviations during interactions, this may result in slower improvement and increased difficulty in acquiring a subtle "native-likeness".

Conclusions: Answers to Three Research Questions

In conclusion, I would like to provide a brief answer to each of the three research questions. All conclusions assumed the levels of non-native learners of Japanese between high-intermediate and low-advanced, i.e., above the level where learners manage a conversation with a native speaker using pre-fabricated grammatical structures, but below the level where learner’s speech involves almost no errors or deviations and sounds like a native speaker.

RQ 1: *Do native speakers’ perceptions agree with grammaticality?* Not always. Although clear grammatical errors sound unnatural to them, such errors do not always affect native speakers’ judgment of “native-like” speech, as long as they do not impede smooth interaction and communication. However, if the proportion of grammar errors and any other kind of deviation is high, it might affect the judgment of “native-like” speech.

RQ 2: *Do native speakers’ perceptions focus on pronunciation?* Not always. It seems that native speakers of Japanese do not care as much about unnatural pronunciation as we might expect. In fact, even inserting an English word with English pronunciation into a Japanese utterance is acceptable as long as the inserted word is comprehensible.

RQ 3: *Are there any other important factors that must be considered?* The most influential factors were those related to interaction and were more significant to native speakers’ perceptions of “native-like” speech rather than linguistic factors, such as the *Lexical/Grammatical Factor* and the *Phonological/Phonetic Factor*. This finding was also supported by correlations of highlighted transcripts and *Ranking; Disfluency* had a more significant correlation with *Ranking* than *Grammar* and *Pronunciation*. The category of *Disfluency* in highlighted transcripts had a close relation to *Fluency Factor* in factor analysis. Therefore, it can be concluded that interaction-related factors are more influential and noteworthy as key elements of “native-like” speech.

In addition to answering these three research questions, the study found from the analysis of *online* and *offline* perceptions that there are two types of unnaturalness: *overt* and *covert*. *Overt unnaturalness* includes obvious errors in grammar, word choice, pronunciation, etc. whereas *covert unnaturalness* is comprised of various kinds of disfluency and is closely related to speech interaction. In the investigation of *online* perception (highlighting unnatural parts on transcripts), native speakers most often picked up clear errors, whereas statistical analyses of *offline* perception (questionnaires and rankings) showed that these errors did not have such a significant influence on native speakers’ overall judgment or impression toward non-native speakers’ speeches.

From these results, I conclude the following: if communicating with native speakers is the learner’s goal, perfect accuracy in grammar and pronunciation should not be the focus for L2 learners. Being too sensitive to such errors may impair smooth interaction or discourage a learner from communicating actively, which

then deprives him/her of the opportunity to acquire *covert naturalness*. Another important aspect is that important linguistic research needs to be validated in interactive settings with non-experts. In a research field that studies people, researchers will need to keep this discrepancy in mind and balance important research findings with non-expert perspectives in their studies.

Acknowledgements

I greatly appreciate all the anonymous reviewers for helpful advice to revise this paper. I also appreciate all the organizers of the first annual conference on language diversity for offering me the opportunity of presenting this study. My special thanks are to the co-editors of this special issue, and especially to, Lauren Mason Carris, who reviewed my paper carefully, took her time to discuss it with me, and gave me a lot of great advice.

Notes

1. The question items were based on the results of a pilot study, Fillmore's (1979) definitions of "fluency," and communication strategies by Dörnyei and Scott (1997). For both the original questionnaire and its English translation, see Ajioka (2008).

2. I used the Statistical Package for the Social Sciences (SPSS) program for statistics in this study. The methods adopted were principal axis factoring and Promax rotation with Kaiser normalization, which is appropriate when the resulting factors are correlated with one another. I determined that there were four factors according to scree plot and Eigenvalues. Furthermore, I excluded eleven variables based on communalities, factor loadings, and interpretability. Considering the purpose of this special issue, which is targeted for the broader audience, I chose not to write technical terms in statistics in the body part of this paper.

3. Naming factors extracted by factor analysis is a researcher's job. I named the four factors based on the communalities of the influential variables on each factor.

4. In this analysis, I excluded 3 participants' highlighted transcripts and ranking scores because some parts of them were incomplete (skipped or completely blank).

5. The value of Ranking indicates the average score of the rankings that the participants gave to each interviewee; Ami's rank is the highest while Dan's is the lowest. The upper values in each cell show the raw number of highlighted parts, i.e., unnatural parts. The values in the lower rows show the percentage to the Total.

6. I employed Spearman's rho to obtain the correlations because Rank is an ordinal scale and the categories do not have a normal distribution.

7. I describe each correlation coefficient following the general convention: (1) 0.0 to 0.2 is regarded as 'very weak or negligible' correlation; (2) 0.2 to 0.4 as 'weak, low (and not very important)' correlation; (3) 0.4 to 0.7 as 'moderate' correlation; (4) 0.7 to 0.9 as 'strong, high' correlation; and (5) 0.9 to 1.0 as 'very strong' correlation.

8. Only clear grammar errors and clearly unnatural pronunciations were classified as these two categories. Those difficult to identify were labeled as another category. Therefore, it is all the more interesting that only clear errors and deviations in grammar and pronunciation amounted to this degree.

References

- Ajioka, M. (2008). *What constitutes “nativelike” Japanese? An analysis of perception of native Japanese speakers*. Unpublished master’s thesis, University of California, Los Angeles.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- Dörnyei, Z. & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning*, 47, 173–210.
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 57-83). San Francisco: Academic Press.
- Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-131). Mahwah, NJ: Erlbaum.
- Flege, J. E., Munro, M., & MacKay, I. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.
- Johnson, J. & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Pawley, A. & Syder, F. H. (1983a). Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics*, 7, 551–579.
- Pawley, A. & Syder, F. H. (1983b). Two puzzles for linguistics theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–227). London: Longman.
- Pawley, A. & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–199). Ann Arbor, MI: The University of Michigan Press.
- Uemura, R. (1998). *Nihongo kaiwa deeta beesu no koochiku to danwa bunseki (The construction of database of Japanese conversation and discourse analysis)*. In the study supported by Grant-in-Aid for Scientific Research on Priority Areas “Humanities and Computer,” the Ministry of Education, 1996–1998.

Mayumi Ajioka is a doctoral student in applied linguistics at UCLA. Her main research interest is native-like fluency among non-native speaker speech and native speaker perceptions in conversing with non-native language speakers.

