Interspeech 2022
18-22 September 2022, Incheon, Korea

Development of allophonic realization until adolescence: A production study of the affricate-fricative variation of /z/ among Japanese children

Sanae Matsui[1,2], *Kyoji Iwamoto*[1,3], *Reiko Mazuka*[1,4]

[1]Riken Center for Brain Science
[2]Sophia University
[3]Waseda University
[4]Duke University
sanae.matsui@riken.jp, kyoji.iwamoto@riken.jp, reiko.mazuka@riken.jp

Abstract

The development of *allophonic variants* of phonemes is poorly understood. Thus, this study aimed to examine when children typically begin to articulate a phoneme with the same allophonic variant typically used by adults. Japanese children aged 5–13 years and adults aged 18–24 years participated in an elicited production task. We analyzed developmental changes in allophonic variation of the phoneme /z/, which is realized variably either as an affricate or a fricative.

The results revealed that children aged nine years or younger realized /z/ as affricate significantly more than 13-year-old and adult speakers. Once the children reached 11 years of age, the difference compared to adults was not statistically significant, which denotes a similar developmental pattern as that of speech motor control (e.g., lip and jaw) and cognitive-linguistic skill. Moreover, we examined whether the developmental changes of allophonic realization of /z/ are due to speech rate and the time to articulate /z/. The results showed that the allophonic realization of /z/ is not affected by these factors, which is not the case in adults. We also found that the effects of speech rate and the time to articulate /z/ on the allophonic realization become adult-like at around 11 years of age.

Index Terms: language development, allophone, manner, affricate

1. Introduction

1.1. Segmental development of speech production

Most early studies on the development of segment production were based on *phonemes,* that is, speech sounds that distinguish words from each other [1]. To our knowledge, there are few studies that analyzed the *allophonic variants of phonemes*, that is, speech sounds that do not distinguish word meanings, in child speech. For example, the English phonemes /t/ and /d/ are produced as not just the canonical forms [t] or [d] but also as three allophonic variants: the unreleased stop [t̚] or [d̚], the flap [ɾ], and the glottal stop [ʔ]. Thus, words like *writer* are often realized with flap sound [ɾ], not [t].

Song et al. [2] showed that 2-year-old English speakers produced canonical variants [t] or [d] of phoneme /t/ or /d/ more than other three allophonic variants, which is not the case in adult speakers. This study showed that children start to produce phonemes as canonical variants rather than the other three allophonic variants. However, it is still unclear when children produce the allophonic realization of phonemes in an adult-like fashion. In the current study, we address this issue. This question is crucial to consider when adult-like speech production is achieved because language is a social tool.

Some finer phonetic characteristics within the phonemic category become like adults in adolescence or later. Vowel space area in child speech is larger than in adults and does not reach that of adults even at the age of 14 years [3]. VOT values for voiced and voiceless stops at 5–9 years are longer and more variable than that at 11–13 years [4]. Child speakers aged 8–12 years produced longer and more variable closure duration and duration of voicing during closure for word initial /b/ compared to adults [5]. These developments seem prolonged, given that children of approximately three years of age show adult-like grammatical knowledge [6,7] These prolonged developments are often ascribed to the development of speech motor control.

1.2. Other aspects of speech development until adolescence

Other than detailed phonetic characteristics of phonemes, development continues into adolescence in several aspects of speech production. Speech rate increases up to 13 years of age [8, 9, 10]. Speech error also decreases with age up to about 11 years [11]. This prolonged development is attributed to the gain of speech motor control as well as those of cognitive and linguistic processing. For example, speech rate increases with age because pause time decreases with age driven by cognitive-linguistic development and articulatory displacement decreases with age driven by that of speech motor control [8].

1.3. Japanese phoneme /z/

In the present study, the voiced phoneme /z/ was chosen as the target segment. In standard Japanese, there are plain mora (or *chokuon* in Japanese) and palatalized mora (or *youon*) and they are regarded as different phonemes because there are minimal pairs between plain consonant [z] and palatalized one [ʒ]. They are realized variably either as an affricate [d͡z][d͡ʒ] or a fricative [z][ʒ], respectively. This study treats both phonemes as target segments. In this paper, /z/ and /ʒ/ are represented as /z/ for simplicity, unless a distinction is required.

1.4. Allophonic variation of /z/

It is important to note that the variations of /z/ are not rule-based categorically conditioned positional allophones [12]. One of the traditional textbooks of Japanese phonetics states that /z/ is a categorically conditioned allophone. Word initial /z/ is realized as affricate and word medial one as fricative [13].

By contrast, Maekawa [12] states that /z/ is a continuous and noncategorical variation rather than categorically

10.21437/Interspeech.2022-346

conditioned allophones. He examined the traditional account using a large-scale corpus of adult Japanese speakers and showed that /z/ does not have categorically positional variants, but that its realization is mostly determined by the *time allotted for consonant articulation* (TACA). TACA is defined as the acoustic duration of /z/ plus that of the preceding pause or closure. Its definition reflects the observation that when we have enough time to articulate /z/, it tends to be realized as an affricate. That is, as TACA increases, /z/ is more likely to be affricate than fricative.

1.5. Aim

This study aimed to examine when children typically begin to articulate the phoneme /z/ in Japanese with the same allophonic variant typically used by adults.

2. Method

Japanese children aged 5–13 years and college-aged adults participated in an elicited production task. The speech data were originally collected for a large-scale project [10]. A brief overview of the recording and the description of the portion of data used in the current analysis are given in this section. Detailed features of the experimental design are referred to [10].

2.1. Participants

Participants were six age groups of 108 Japanese native speakers (18 each of 5-, 7-, 9-, 11-, 13-year-old children, and adults). Adult speakers were aged 18–24 years (mean= 20.11 years). None of the participants had any known speech or hearing difficulties.

2.2. Stimuli

Stimuli were six nonwords with word initial /z/ and /ʒ/: /zana/, /za:na/, /zanara/, /ʒana/, /ʒa:na/, and /ʒanara/. All words were accented on the first mora. These nonwords were embedded in three types of carrier sentences that varied in the length. A short carrier phrase was /__ dajo/ “This is __.” A middle length one was /X no __ dajo/ “This is __ with X,” in which X represents nine noun words that were three moras in length and accented on the first mora (e.g., /tómato/ “tomato”). A long one was /X no __ wa Y ni irujo/ “__ with X is in Y,” where X was a noun as mentioned above, and Y was one of six nouns that were three or four moras in length (e.g., /sunaba/ “sandpit”). Thus, target consonants were either sentence-initial or after the vowel /o/. The examples are shown below.

(Short) /Zana dajo/ “This is Zana.”

(Middle) /Tomato-no zana dajo/ “This is Zana with a tomato.”

(Long) /Tomato-no zana-wa sunaba ni irujo/ “Zana with a tomato is near the sandpit.”

2.3. Procedure

An elicited production task was conducted in a quiet room. A female Japanese-monolingual experimenter and a participant sat across a table facing each other with a head-mounted condenser microphone (SHURE BETA54) connected to a recorder (Marantz PMD661MK). Six different toys were used, with each assigned a target word. The experiment was divided into three blocks. In the first block, the experimenter showed the participants a picture of a stuffed animal-like character and told them its name. Then, the participants were asked to remember and say that name in a short sentence at least twice (e.g., /zana dajo/ “This is Zana”). Their production was preceded by the question “Who is this?”

In the second block, the experimenter showed two pictures of the same character, which carried an item such as tomato. The participants were asked to say the name in a middle-length sentence (e.g., /tomato no zana dajo/ “This is Zana with a tomato”) after the question “Who is this?”

In the third block, a large picture of a park was placed on a table, wherein there were eight different playground equipment (six of the equipment names were used in long-carrier sentences mentioned in 2.2 as Y.) The experimenter placed a character holding an item onto two of the playgrounds, and then asked the participants to say a long sentence (e.g., /tomato no zana ga sunabani irujo/ “Zana with a tomato is in the sandpit”). They produced the target words after the question “Where is this?”

We elicited each target word at least six times correctly (short sentence: 2; middle sentence: 2; long sentence: 2). The experimenter produced target words only when she told the participants the character's name so that they would not imitate her pronunciation. The order of target words and combinations between the toy characters and the items, and where each picture was placed in the park were counter-balanced across participants and target words. Each session lasted 15–35 minutes. These experimental blocks were preceded by a practice block with two characters not used in the experiment so that the participants could remember charier phrases.

2.4. Coding and measurement

A total of 5,480 tokens were annotated using Praat [14] by six Japanese-native-speaking coders with phonetics training and checked by the first author. The dependent variable of this study is a binary variable representing the manner of articulation of /z/. The affricate and fricative were distinguished by the presence or absence of a burst-like noise (Fig. 1). Additionally, TACA was measured following the definition of [12]. TACA is defined as the acoustic duration of /z/ plus that of the preceding pause or closure. However, if the duration of the preceding pause or closure is longer than that of /z/, TACA is defined as the duration of the /z/ multiplied by two. Thus, TACAs of /z/ in the short phrase were always the duration of the /z/ multiplied by two since /z/ in the short phrase is sentence initial.

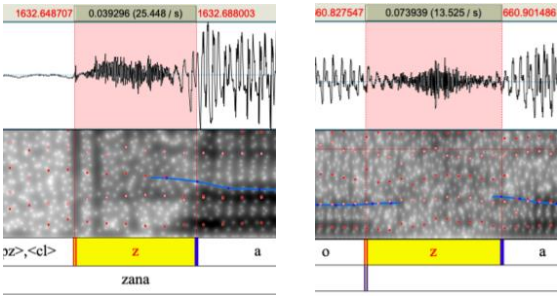

Fig 1: *Affricate (left) and fricative (right) realizations of /z/.*

Sentences in which the coders observed inaccuracies or pauses were excluded from the current analysis [10]. Among the total of 1,431 instances of /z/ and 1,498 of /ʒ/ used in the current analysis, 948 and 1,218 were realized as affricates, respectively. Therefore, the overall RAA (rate of affricate articulation) was 66.2 % for /z/ and 81.3% for /ʒ/.

3. Results

3.1. Rate of Affricate Articulation (RAA) by age

Figure 2 shows the RAA of /z/ and /ʒ/ by each age group. Each blue dot represents a speaker.

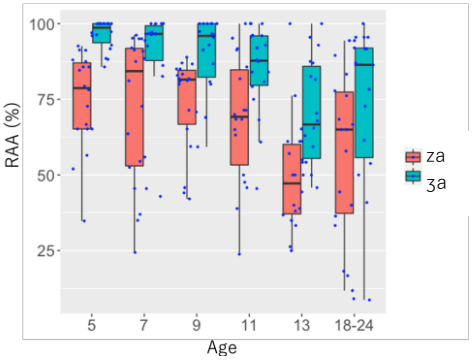

Fig 2: *RAA of /z/ and /ʒ/ by age groups.*

To test whether and how the manner of articulation of /z/ changes with age, we constructed a generalized linear mixed-effects model using *glmer* function included in the *lme4* package [15] for the *R* (version 4.1.2) [16]. The model used manner as the binary dependent variable where 1 represents affricate and 0 represents fricative. The model contains age, consonant (/z/ or /ʒ/), and an interaction term between them as predictors, and speaker and word as random intercepts to ensure that the model generalizes to other speakers and items. The results of this model are presented in Table 1.

The main effect of age was significant, indicating that /z/ is more likely to be a fricative than affricate as the age increases. Furthermore, the significant main effect of the consonant (ʒ) was found, indicating that /ʒ/ is more likely to be an affricate compared to /z/. The significant interaction was also found.

Table 1: *Results of the logistic regression model for manner of articulation, with age and consonant*

Model1	Coefficient	s.e.	*P*	
Intercept	1.438	0.334	<.0001	***
Age	-0.064	0.024	<.001	***
Consonant (ʒ)	1.841	0.310	<.0001	***
Age: Consonant (ʒ)	-0.056	0.015	<.0001	***

Multiple comparisons among age groups with Tukey corrections were performed using *emmeans* function [17]. The results are summarized in Table 2, demonstrating that children aged nine years or younger realized /z/ as affricate significantly more than 13-year-old and adult speakers. Once the children reached 11 years of age, the difference compared to adults was not statistically significant.

Table 2: *The p-values of multiple comparisons among age groups. Significance level is indicated by *p<.05, **p<.01, ***p<.001*

Age	**7**	**9**	**11**	**13**	**Adults**
5	0.96	0.84	0.31	<.001 ***	<.05 **
7	-	0.99	0.82	0.0051 **	0.022 *
9	-	0.95	0.015 *	0.057
11		-	0.16	0.37
13			-	0.99

Interestingly, this developmental pattern is consistent with that of speech rate and frequency of speech error [8-11]. As discussed above, these developmental changes are due to cognitive-linguistic processing and speech motor control [8]. The age-related changes in the realization of /z/ may be related to gains in cognitive, linguistic, and speech motor skills.

3.2. Contribution of speech rate and TACA to the RAA

We also examined whether speech rate and time to articulate /z/ made significant contributions to the manner of articulation of /z/ above and beyond the age factor. Given that the developmental pattern of /z/ realization is similar to that of speech rate and that /z/ tends to be a fricative as speech rate increases [12], it seems that /z/ is likely to be a fricative with age since speech rate increases with age. In addition, the pause duration of adult speakers is shorter than that of child speakers [8]. Meanwhile, the phonetic realization of /z/ is sensitive to pause duration since the manner of articulation of /z/ is mostly determined by whether the time to articulate /z/ is sufficient (see 1.4.). Thus, it appears that /z/ is likely to be a fricative with age since the time to articulate /z/ shortens with age.

We examined whether speech rate and TACA made a significant contribution to the manner of articulation above the age factor. We built a generalized linear mixed-effect model, which used affricate and fricative coded as 1 and 0, respectively, as the dependent variable. The model included age, speech rate, TACA, consonant (/z/ or /ʒ/), an interaction term between age and speech rate, and between age and TACA as fixed factors, and speaker and word as random intercepts. Speech rate was computed for each sentence by dividing the total number of morae in the sentence by the total duration time (in milliseconds) of the sentence, resulting in mora per millisecond. A Pearson correlation coefficient was computed to confirm whether speech rate and TACA were correlated. There was a slight correlation between them: r(2916)=.20, *p*<.0001. Speech rate and TACA in the model were standardized within each variable, not for each participant. Table 3 shows the estimated parameters for the model.

Table 3: *Results of the logistic model for manner of articulation, with age, speech rate, TACA, and consonant*

Model2	Coefficient	s.e.	*p*	
Intercept	-0.146	0.495	<.05	***
Age	0.001	0.040	0.822	n.s.
Speech rate	-0.064	0.251	0.796	n.s.
TACA	-0.393	0.160	<.05	*
Consonant(ʒ)	1.281	0.249	<.001	***
Age:Speech rate	0.093	0.025	<.0001	***
Age:TACA	0.091	0.014	<.0001	***

The main effects of age and speech rate were not significant. The main effect of TACA was significant, indicating that /z/ is more likely to be a fricative as TACA becomes longer, which is inconsistent with the findings of Maekawa [12]. Significant interactions between age and speech rate as well as age and TACA were found.

Figure 3 shows the probabilities of affricate articulation predicted by the generalized mixed model. The x-axis is standardized values of speech rate (left) and those of TACA (right). The y-axis is the probabilities of affricate articulation since the dependent variable of the model was coded affricate as 1 and fricative as 0. The left figure shows that the probabilities of affricate articulation increase as speech rate becomes slower in children aged nine years or older. In contrast, those of younger speakers remain flat, indicating that the effect of speech rate is small in 5- to 9-year-old groups. Similarly, the right figure shows that the probabilities of affricate articulation increase as TACA become longer in children aged 11 years or older, replicating Maekawa [12]. However, the effect of TACA is small in 5- to 9-year-olds.

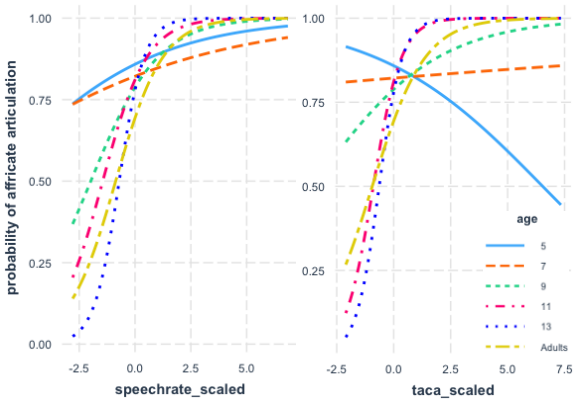

Fig 3: *Effects of speech rate and TACA on probability of affricate articulation by age groups.*

4. Discussion

4.1. Development of phonetic realization of /z/

In this production study, we explored when allophonic variants of phonemes are realized in an adult-like fashion, with respect to Japanese children and adults' production of /z/.

We found that, /z/ is more likely to be a fricative with age. This change occurs until around 13 years, and children aged 11 years achieved adultlike allophonic realization of /z/. This developmental pattern is similar to that of speech rate and pause duration. Then, we hypothesized that affricate articulation decreases with age since speech rate increases and time to articulate /z/ (defined as TACA) shortens with age.

We examined whether speech rate and the time to articulate /z/ contribute to the manner above the age factor. The results revealed that those factors may not directly contribute to the age-related changes of /z/ realization, because the main effects of those factors were not significant. Rather, the effects of speech rate and the time to articulate /z/ changed with age. Children aged 11 years or older showed adult-like manner of alternation, that is /z/ tends to be an affricate when speech rate is slow and the time to articulate /z/ is sufficient. In contrast, 5- to 9-year-olds showed small effects of speech rate and TACA.

Given that development of /z/ showed a similar pattern to that of cognitive, linguistic, and speech motor control skills, developmental changes of /z/ might be due to these skills. For example, Nip and Green [8] showed that articulatory displacement is large in childhood and decreases with age up to 13 years. It appears that children produce /z/ as an affricate regardless of speech rate since their articulation control is immature and their tongue movement is large, meaning the tips of their tongues reach the alveolar ridge more often than adult speakers.

4.2. Allophonic variation and finer phonetic characteristics

This study treated /z/ as a target segment, which might not be categorically conditioned allophones, and indicated that adult-like allophonic realization is achieved in adolescence. Previous studies also showed that detailed phonetic properties (e.g., vowel space area, VOT, or closure duration of stops) develop up to adolescence (see 1.1.). However, the variation of /z/ differs from the other three detailed phonetic variations in that /z/ is a language specific variation [18,19]. While the variation of vowels, VOT, and closure duration of stops within a phonemic category might be found in several languages, the affricate-fricative variation of /z/ is limited in world languages; however, Japanese children must learn them. Further research is needed to examine whether the development of /z/ and that of other phonetic details are driven by same mechanisms, and to specify which factor drives the development of /z/.

5. Conclusions

This study provides the first investigation of the allophonic realization of /z/ in child and adult speech. We found that the allophonic variation in terms of manner of articulation develops until around the age of 13 years, and children aged 11 years achieve adult-like phonetic realization. Moreover, the effects of speech rate and the time to articulate /z/ on the manner of articulation also become adult-like at around the age of 11 years. Given that the developmental pattern of /z/ is similar to that of cognitive, linguistic, and speech motor control skills, those skills may contribute to the development of allophonic realization of /z/.

6. Acknowledgements

This study was funded in part by the RIKEN Junior Associate Program, JSPS Grant-in-Aid for Scientific Research S (16H06319), Specially Promoted Research (20H05617), MEXT Grant-in-aid for Innovative Areas #4903 (17H06382), and NINJAL collaborative research project "Cross-linguistic Studies of Japanese Prosody and Grammar." The ethics approvals for this study were Wako3 24–11(13)–(16). We are grateful to Dr. Akiko Hayashi and Ayako Kondo at Tokyo Gakugei University, from which all participants were recruited and tested. We would also like to thank Ken'ya Nishikawa, Kakeru Yazawa, and Hyun Kyung Hwang, who assisted in transcribing and coding the recordings, and Dr. Mafuyu Kitahara, who commented on the manuscript. We also thank the children and adults who participated in this study.

7. References

[1] M. Ota. L1 Phonology: Phonological Development. *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*. Berlin: De Gruyter Mouton, 2015.

[2] J. Y. Song, S. Shattuck-Hufnagel, and K. Demuth. "Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop/t, d/codas," *Journal of Phonetics*, vol. 52, pp. 152–169, 2015.

[3] M. Pettinato, O. Tuomainen, S. Granlund, and V. Hazan. "Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication," *Journal of Phonetics*, vol. 54, pp. 1–14, 2016.

[4] S. P. Whiteside, R. Dobbin, and L. Henry. “Patterns of variability in voice onset time: A developmental study of motor speech skills in humans,” *Neuroscience Letters*, vol. 347 no. 1, pp. 29–32, 2003.
[5] E. Jacewicz, L. J. Arzbecker, R. A. Fox, and S. Liu. “Variability in within-category implementation of stop consonant voicing in American English-speaking children,” *The Journal of the Acoustical Society of America*, vol. 150, no. 5, pp. 3711–3729, 2021.
[6] E. Hoff. *Language Development*, 5th Edition. Bel-mount, CA: Wodsworth, 2003.
[7] B. Lust. *Child Language: Acquisition and Growth*. Cambridge: Cambridge University Press, 2006.
[8] I. S. Nip and J. R. Green. “Increases in cognitive and linguistic processing primarily account for increases in speaking rate with age,” *Child Development*, vol. 84, no. 4, pp. 1324–1337, 2013.
[9] O. Amir and D. Grinfeld. “Articulation rate in childhood and adolescence: Hebrew speakers,” *Language and Speech*, vol. 54, no. 2, pp. 225–240, 2011.
[10] I. Kyoji, K. Hideaki, and M. Reiko “Speech rate development in Japanese-speaking children and proficiency in Mora-timed rhythm,” *Journal of Experimental Child Psychology* (In Press), 2022.
[11] K. Iwamoto, H. Kikuchi, and R. Mazuka, “Children’s speech inaccuracies and developmental change: An elicited production study in 5- to 13-year-old Japanese children,” *Journal of the Phonetic Society of Japan*, vol. 25, pp. 67–86, 2021.
[12] K. Maekawa. “Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of/z/in spontaneous Japanese,” *Journal of Phonetics*, vol. 38, no. 3, pp. 360–374, 2010.
[13] Y. Amanuma, K. Otsubo, and O. Mizutani, O. *Nihongo Onseigaku* (Japanese phonetics). Tokyo: Kuroshio Publishing, 1982.
[14] P. Boersma, and D. Weenink. “Praat: Doing phonetics by computer. Version 6.1.54,” 2021. https://www.fon.hum.uva.nl/praat/.
[15] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823*. 2014.
[16] A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen. “lmerTest package: Tests in linear mixed effects models,” *Journal of Statistical Software*, vol. 82, pp. 1–26, 2017.
[17] R. Lenth, H. Singmann, J. Love, P. Buerkner, and M. Herve. “Package ‘emmeans’,” 2019. https://cran.r-project.org/web/packages/emmeans/index.html.
[18] K. P. Mohanan. *The theory of lexical phonology (Vol.6)*. Springer Science & Business Media, 2012.
[19] A.C. Cohn. “Nasalisation in English: phonology or phonetics,” *Phonology*, vol. 10, no. 1, pp. 43–81, 1993.