The impact of age on the naturalness of /r/ productions

McNair Research Paper

Marissa LeDoux

Faculty Advisor:
Dr. Breanna Krueger

Graduate Student Mentor:
Josephine Cox

Department of Communication Disorders
McNair Scholars Program, University of Wyoming

1 June 2021

Abstract

This study investigated the influence of speech therapy introduced in early childhood versus introduced later in childhood. Specifically, this study compared children who started speech therapy for the production of the /r/ sound at age 4-5 as compared to those who started at age 7-8. This research included 2 types of measures: objective acoustic measures and subjective perceptual judgments. First, naturalness was assessed objectively by analyzing the acoustic properties of the /r/ sound (e.g. F2-F3 distance and duration). Second, recruited adult participants were asked to make judgments about the naturalness of the /r/ production of these recordings of younger and older children producing the /r/ sound before and after treatment. The participants were asked to rate each production on a scale of 1 to 7, with 1 representing highly unnatural sounding speech and 7 representing highly natural sounding speech. Finally, the acoustic measures and perceptual judgments were examined side by side to see if they correlated with one another. The results of this study suggest that it is better for children to start speech therapy at an earlier age in childhood (ages 4-5) in order to have the best outcome for naturalness, in terms of acoustic measurements and perceptual judgments. These findings are the start to helping speech-language pathologists determine the right age to enroll a child in speech therapy for the /r/ sound.

Introduction

The early childhood years are important to speech sound development and often determine whether a child is on the path to typical speech sound development. With the phoneme /r/ being one of the last sounds to emerge in children, (McLeod & Crowe, 2018; Crowe & McLeod, 2020) it also tends to be a common error sound treated in speech therapy (Klein et al., 2012). Past developmental norms have stated that the /r/ sound is not developed until age 6

(Sander, 1972), however, the new norms state that children develop the /r/ sound earlier than this, around age 4-5 (McLeod & Crowe, 2018). Due to the fact that current developmental norms differ, speech-language pathologists have a difficult time determining the right time to enroll a child with speech language impairments into speech therapy. This is particularly true in sounds where there is disagreement between the norms, such as the /r/ sound. Some traditional approaches in determining when to enroll a child into speech therapy include; referring to the developmental norms, using normative data about error types and patterns, and using standardized tests, such as the Goldman-Fristoe Test of Articulation (Storkel, 2019). Combining various measures, such as the ones listed above, is suggested for determining a good time to enroll a child into speech therapy.

Naturalness of Speech

Naturalness is a measure that is used in the field of speech-language pathology. As stated in a study done by Klopfenstein and colleagues (2020), naturalness is an overall measure of an individual's speech that sounds natural to the listener. Naturalness is often closely related to the individual's prosody and intelligibility (Klopfenstein et al., 2020). Prosody refers to an individual's intonation, stress pattern, loudness, pausing, and rhythm. Whereas, intelligibility refers to how well the speaker is understood by the listener. There are two main measurements that typically determine the naturalness of speech, the physical acoustic properties of the sounds and perceptual judgments by listeners. Often, these two measures tend to correlate with one another.

Acoustics of /r/

Acoustic properties, such as F2-F3 distance, have been shown to be a good indicator in separating children with speech-sound disorders to those with normal speech development (Klein et al., 2012). F2-F3 distance refers to the spacing between the second and third formants. Formants are areas of concentrated acoustic energy, which can be observed on spectrograms in both vowels and some consonants that have vowel-like qualities, such as /j/ and /r/. In American English, there are a few things that help us distinguish the /r/ sound on a spectrogram. First, the /r/ sound is typically correlated with a low height of the third formant frequency. Second, the second and third formants for the /r/ sound are relatively close together and sometimes tend to merge (Klein et al., 2012). As a child ages, the mastery of the /r/ sound improves and we see the height and F2-F3 distance decrease (Klein et al., 2012).

Perception of /r/

Children who have difficulty producing the /r/ sound are often described as using /w/ as a substitution (Klein et al., 2012). Klein et al., (2012) mentions that instead of producing a phonemic substitution, the child is actually producing a sound that is acoustically intermediate between /r/ and /w/. “We must move from listener-oriented data to speaker-oriented data in order to reveal what the speaker has achieved acoustically and articulatory” (Scobbie et al., 1996). As adults, we have a more matured phonological and speech motor system, whereas, children do not achieve adult-like phonetic skills until the ages of 8-12 (Scobbie et al., 1996). “Immature speech production routines may be to blame for failing to make phonological distinctions audible to the listener” (Scobbie et al., 1996). Children who use acoustically non-standard cues to mark a distinction between /r/ and /w/ are thought to be closer to achieving accurate productions of the /r/ sound compared to those who use phonemic substitutions (Klein et al., 2012). Sharf and Ohde (1983), conducted a study using synthesized speech to determine the role F2 and F3 plays in determining the “goodness” of /r/ and /w/ sound productions. They found that the ratings of the speech were influenced by the manipulation of the second and third formants. A production was more likely to be rated as a good /r/ as F2 increased. However, when F3 increased, the production was more likely to be rated as a good /w/ (Sharf & Ohde, 1983).

Age and Treatment

Approximately 3.5% to 5% of 4-year-olds have a developmental speech disorder (Morgan et al., 2017). Current practices often involve the system of "delayed or disordered" when referring to young children with speech disorders. Morgan et al., (2017) states that, a delay is an error that is typical for speech sound development, but is delayed in comparison to 90% of children of the same age. When a child has a speech disorder, their speech contains atypical errors that are not typically seen in children at any age during speech development (Morgan et al., 2017). Determining if a child's speech is delayed or disordered is especially important in implementing a plan for treatment. Morgan et al., (2017) found that 40% of children with speech errors at age 4 had those same errors at age 7. The findings of this study proves that speech errors do not resolve for all children by age 5, as previously thought.

Currently, there has not been extensive research on when to begin treatment for children with these types of errors and whether young children (age 4-5) are ready to begin treatment for the /r/ sound. On the other hand, some argue that waiting too long runs the risk of the errors becoming persistent (Shriberg, 1980). The study done by Klein and colleagues (2012) looked at the relationship between listener's perceptions and acoustic properties. The children focused on in their study were between the ages of 6 and 7 years. This study did not look at children pre and post treatment, but rather children with disordered phonology vs. children with age-appropriate phonology. With new data sets suggesting that children develop the /r/ sound at an earlier age than previously thought, this study might not give us the full picture on how the age that a child starts treatment at can affect the naturalness of speech production. Therefore, it was our aim to fill in this gap of knowledge.

Purpose and Research Questions

In the study done by Klein and colleagues (2012), acoustic properties of the /r/ sound were compared against the perception of the speech productions, however, they did not vary age pre and post treatment. Therefore, this study will consider age differences pre and post treatment, in order to fill that knowledge gap.

The purpose of this study included three separate portions. First, we were interested in the impacts of naturalness between the two age groups, in terms of acoustic measurements and perceptual judgments, separately. Second, we were interested in examining the relationship between the acoustic properties of /r/ productions and the perceptual judgements by research participants between both age groups. To examine this relationship, we compared the listeners ratings against the F2-F3 distance for each production. The following research questions were investigated:

1. Does the age that a child starts speech therapy have an impact on naturalness of the production of the /r/ sound in terms of acoustic measures?
2. Do younger or older children have more natural productions of the /r/ sound after speech therapy as judged by adult listeners?
3. How do objective (acoustic) and subjective (perceptual judgments) judgments of /r/ productions for younger and older children correlate after speech therapy?

The purpose of this research was to investigate the influence of age on naturalness of speech sound production by comparing acoustic measurements to perceptual judgments of /r/ sound productions in two different age groups.

It is our aim to examine the influence of a child's age on their treatment outcomes for the /r/ sound in terms of naturalness. Acoustic measures and perceptual judgments of speech have been found to be closely correlated to one another, and will serve as the primary measure for naturalness of the /r/ productions by these children (Klein et al., 2012).

Methods

Participants

The participants included 20 adult listeners, ages 19 to 73. All participants have no history of hearing loss, are monolingual English speakers, and are inexperienced listeners.

Recorded speech samples were drawn from 8 children and included two different age groups. The first group included 3 children who were between the ages of 4 and 5 and the second group included 5 of children who were between the ages of 7 and 8.

Acoustic Properties

Using the software Praat, (Boersma & Weenink, 2021) words containing the /r/ sound in each speech sample were extracted to yield 248 tokens. The tokens contained the /r/ sound in the initial, medial, and final positions. In each token, the duration for the /r/ sound was noted. Then, wideband spectrograms of each production were visually inspected with a view range of 0-6000 Hz. Praat's software provides automated linear predictive coding format tracking, in which red dotted lines represent formants 1-5. Using this, we selected a point within the third formant from which to extract a window for formant measurements. A 14 ms window was then extracted around the selected point. The formant values were then automatically computed by Praat and formant heights of F2 and F3 were recorded in a spreadsheet. The distance between F2 and F3 were then calculated by subtracting the F3 value from the F2 value.

Reliability checks of the acoustic properties of the tokens were implemented. The duration and F2-F3 distance of 51 tokens, (20% of the full sample) were also measured by a graduate student. To ensure that the first measurements were accurate, a t-test was implemented for both duration and F2-F3 distance. The t-test revealed no significant difference in terms of duration [$t(-0.07)=101.91$, $p=0.470$) or F2-F3 distance [$t(0.12)=100.86$, $p=0.454$] as measured by the undergraduate and graduate student.

Perceptual Judgments

248 tokens were uploaded into the software, Gorilla (Anwyl-Irvine, et al., 2019). Gorilla is a software that allows researchers and students to create and perform experiments online. Gorilla played each token one-by-one to the participants. Each participant was then asked to rate the token's naturalness on a scale of 1 to 7; 1 being highly unnatural sounding speech and 7 being highly natural sounding speech. To select a rating of each production, the participants would select a number on the screen and Gorilla would then move them onto the next token. This process was repeated until the participants listened to and rated all 248 tokens.

Results

Acoustic Analysis

Averages of F2-F3 distance were taken for both groups at pre-test and post-test. As shown in Table 1, the average F2-F3 distance for younger children at post-test was much lower than the older children's F2-F3 distance at post-test. This suggests that acoustically, the younger children had better /r/ productions at post-test. Table 1 also shows that the younger children decreased their F2-F3 distance from pre-test to post-test more than the older children. The younger children decreased their F2-F3 distance by 85.004 Hz, whereas, the older children only decreased their F2-F3 distance by 9.782 Hz. This suggests that younger children succeeded more at improving their /r/ productions than the older children, in terms of acoustic measures. Next, we ran a t-test to determine if there was a significant difference between the average F2-F3 distances from pre-test to post-test. For the younger kids, we were given the following results: [t(0.77)=42, p=0.446]. For the older kids, we got the following results: [t(0.14) =73, p=0.889]. With both of the p-values greater than 0.05, the results were insignificant. This means that the difference between the average F2-F3 distances from pre-test to post-test were not significantly different. We believe this to be insignificant because some tokens at post-test were still inaccurate productions. If we got rid of the inaccurate tokens at post-test and looked only at the accurate tokens, we would expect to see these t-tests to yield a significant difference value.

Table 1. F2-F3 Distance Averages Pre/Post Test Results

Age Group	Average Pre-Test	Average Post-Test	Pre/Post T-Test
Young Children	1350.865 Hz	1265.861 Hz	[t(42)=0.77, p=0.446]
Old Children	1649.623 Hz	1639.841 Hz	[t(73)=0.14, p=0.889]

Next, we ran a t-test for younger vs. older children at posttest only. This yielded a significant difference of: [t(-4)=114, p<0.0001]. At first, we believed this significant difference could be due to a pitch difference, however, when looking at the ranges of the two age groups, which are shown in Table 2, they overlap. Therefore, pitch was likely not the cause of this significant difference. The younger children improving their /r/ productions more than the older children was the cause of this significant difference.

Table 2. Pre/Post-Test Min/Max Data

Age Group	Pre-Test Min	Pre-Test Max	Post-Test Min	Post-Test Max
Young Children	278.1547 Hz	2813.99 Hz	219.8562 Hz	2419.229 Hz

Old Children	171.2861 Hz	2705.637 Hz	261.8681 Hz	2620.269 Hz

Perceptual Judgements

When looking at the data for the perceptual judgements portion of this study, we ran two tests; the Mann-Whitney U test and the Spearman Correlation test. We chose to do the Mann-Whitney U test-because age is a nominal type of data and the ratings are an ordinal type of data. The Spearman Correlation test was chosen because the ratings are an ordinal type of data and F2-F3 distance is an interval type of data. When running the Mann-Whitney U test for the younger children, the results yielded these values (W= 7964, p =0.017). This means that there is a significant difference between the response value at pre-test vs. at post-test. Adult participants rated children's /r/ productions more accurate at post-test than at pre-test for younger children. Table 3 shows the average adult ratings for both age groups at pre-test and post-test. When running the Spearman Correlation test for the younger children, the results yielded these values: (r=.009, p= 0.76). Therefore, there was no significant association between ratings of speech sound production and F2-F3 distance overall.

Table 3. Pre-Test and Post-Test Adult Response Averages and Standard Deviations

Age Group	Pre-Test	Post-Test
Young Children	M=3.37 SD=1.87	M=3.61 SD=1.81
Old Children	M=3.62 SD=1.90	M=3.65 SD=1.85

As shown in Table 4, the Mann-Whitney U test for older children yielded the result values of (W= 536720, p =0.629). Therefore, there was no significant difference between the response value at pre-test vs. post-test. When running the Spearman Correlation test for the older children, the results yielded these values: ($r= -0.04147424$, $p= 0.059$). This means there was a correlation between ratings of speech sound production and F2-F3 distance and our data was trending.

Table 4. Perceptual Judgements Results

Age Group	Mann-Whitney U	Spearman Correlation
Young Children	($W= 7964$, $p =0.017$)	($r=.009$, $p= 0.76$)
Old Children	($W= 536720$, $p =0.6296$)	($r= -0.04147424$, $p= 0.05983$)

Discussion

Krueger (2017), found that younger children ages 4-5 are just as good at learning the /r/ sound in treatment as compared to older children ages 7-8. However, due to the differing norms across different data sets, young children are not typically enrolled in speech therapy until a later age, for late acquired sounds, such as /r/. The purpose of this study was to investigate the influence of speech therapy by comparing children who started speech therapy for the production of the /r/ sound at age 4-5 as compared to those who started at age 7-8. The results of this study suggest that it is better for children to start speech therapy at an earlier age in order to get the best results in terms of naturalness, acoustically and perceptually. The younger children in this study, (ages 4-5) decreased their F2-F3 distance more at post-test, than the older children did (ages 7-8). Adults in this study rated the younger children's /r/ productions more accurate at post-test than pre-test.

With different data sets suggesting that children develop sounds at different ages, clinicians may wait until the child is older to enroll them into speech therapy. However, with the findings of Krueger (2017), and the findings of the present study, it would be beneficial for clinicians to enroll a child into speech therapy at an earlier age. Younger children are capable of learning sounds as quickly as older children and have a better outcome from speech therapy, acoustically and perceptually.

Limitations

The present study looked at all tokens pre-test and post-test to gain a look at the whole picture. Therefore, some tokens, at post-test still contained inaccurate /r/ sound productions. This may have caused our F2-F3 distance t-test results to yield an insignificant value. It is possible that when only looking at accurate tokens at post-test that these t-test results would then yield a significant value.

Future Directions

In a future direction for this study, the limitation listed above would be addressed. We would get rid of the inaccurate /r/ sound productions at post-test, and focus only on the accurate ones. This would give us a better picture on if there is a significant difference in acoustic measurements between the F2-F3 distance pre-test versus post-test, for both age groups.

Conclusions

The present study examined how starting speech therapy earlier in childhood (ages 4-5) versus later in childhood (ages 7-8) has an impact on the naturalness of /r/ sound productions. This study found that younger children had more natural sounding /r/ sound productions after treatment in terms of acoustic measures. The younger children also had more natural /r/ sound productions after speech therapy as judged by adult listeners. Acoustically and perceptually, younger children had more natural sounding /r/ productions, so the two measures did correlate with one another. The findings of this study suggests that it is better for children to start speech therapy at an earlier age in childhood, in order to have the best results in terms of naturalness.

These findings can be used to guide speech-language pathologists to implement a more effective plan and timeline for speech therapy.

Acknowledgements

Thank you to all of the individuals who participated in this study. I would also like to thank the McNair Scholars program at the University of Wyoming for helping me conduct this research. Without this program's guidance and support, I would not have the opportunity to conduct this study. Thank you to my graduate student mentor, Josie Cox, for all of your help and support. Finally, I extend my utmost appreciation to Dr. Breanna Krueger. Without your help this study would not have been possible. You have pushed me to find a new side of myself as a student that I would have never thought was possible.

References

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods.* https://doi.org/10.3758/s13428-019-01237-x

Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.52, retrieved 28 August 2021 from http://www.praat.org/

Crowe, K., & McLeod, S. (2020). *Children's English Consonant Acquisition in the United States: A Review | American Journal of Speech-Language Pathology*. (2020). Retrieved June 24, 2021, from https://pubs.asha.org/doi/10.1044/2020_AJSLP-19-00168

Klein, H. B., Grigos, M. I., Byun, T. M., & Davidson, L. (2012). The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. Clinical Linguistics & Phonetics, 26(7), 628-645. https://doi.org/10.3109/02699206.2012.682695

Klopfenstein, M., Bernard, K., & Heyman, C. (2020). The study of speech naturalness in communication disorders: A systematic review of the literature. *Clinical Linguistics & Phonetics*, *34*(4), 327–338. https://doi.org/10.1080/02699206.2019.1652692

Krueger, B. I. (2017). *Age as a Factor in the Treatment of Late-Acquired Sounds*. https://kuscholarworks.ku.edu/handle/1808/26338

McLeod, S., & Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. American Journal of Speech-Language Pathology, 27(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

Morgan, A., Ttofari Eecen, K., Pezic, A., Brommeyer, K., Mei, C., Eadie, P., Reilly, S., & Dodd, B. (2017). Who to Refer for Speech Therapy at 4 Years of Age Versus Who to "Watch and Wait"? *The Journal of Pediatrics*, *185*, 200-204.e1. https://doi.org/10.1016/j.jpeds.2017.02.059

Sander, Eric K. (1972). When are Speech Sounds Learned? Journal of Speech and Hearing Disorders, 37 (1), 55-63. https://doi.org/10.1044/jshd.3701.55

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). *Covert contrast as a stage in the acquisition of phonetics and phonology: Working paper*. https://eresearch.qmu.ac.uk/handle/20.500.12289/10

Sharf, D. J., & Ohde, R. N. (1983). Perception of distorted "R" sounds in the synthesized speech of children and adults. *Journal of Speech & Hearing Research, 26*(4), 516–524. https://doi.org/10.1044/jshr.2604.516

Shriberg, L. D. (1980). An intervention procedure for children with persistent/r/errors. Language, Speech, and Hearing Services in Schools, 11(2), 102-110

Storkel, H. L. (2019). Using Developmental Norms for Speech Sounds as a Means of Determining Treatment Eligibility in Schools. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 67–75. https://doi.org/10.1044/2018_PERS-SIG1-2018-0014