Voice assistant technology continues to underperform on children’s speech

Holly Bradley,1 Madeleine E. Yu,1,2 and Elizabeth K. Johnson1,2

1Department of Psychology, University of Toronto, ON M5S 1A1, Canada; holly.bradley@utoronto.ca

*2Department of Psychology, University of Toronto, Mississauga, ON L5L 1C6, Canada;*

*madeleine.yu@mail.utoronto.ca, elizabeth.johnson@utoronto.ca*

**Abstract:** Voice Assistant (VA) technology is increasingly part of children’s everyday lives. But how well do these systems understand children? No study has asked this with children under 5 years. Here, two versions of Siri, and one of Alexa, were tested on their ability to transcribe utterances produced by two, three, and five-year-olds. Human listeners (mothers and undergraduates) were also tested. Results showed that while Siri’s performance on children’s speech has improved in recent years, even the newest Siri and Alexa models struggle with children’s speech. Human listeners far outperformed VA systems with all ages, especially with the youngest children’s speech.

*Keywords:* voice assistant technology, speech perception, speech intelligibility, child language development

1. Introduction

Voice Assistant (VA) Technology has shown dramatic improvements over the past two decades, largely due to advancements in artificial speech recognition (ASR) and machine learning. Speech recognition systems in the early 2000s often had poor hit rates in real-world, noisy environments (Basak et al., 2022). But things are changing. According to Google’s own assessments, their ASR achieved an accuracy rate of around 95% for English speech by 2017, up from around 75% a few years earlier (Proksch et al., 2019). Similarly, Microsoft reported that their ASR technology achieved a 5.1% error rate in the Switchboard dataset (a benchmark for conversational speech) by 2019, comparable to the average human error rate (Kuhn et al., 2024). Current-generation VAs utilize advanced models and self-supervised learning, which have further reduced the error rate. For some languages, it is claimed that these systems now can achieve hit rates at or above human listener accuracy (Kim et al., 2024).

As technology advances, it becomes increasingly woven into our daily lives, shaping how we interact with the world. Siri, the first virtual assistant (VA), launched in 2011 and quickly gained a wide user base, followed by Amazon’s Alexa in 2014. With these developments, technology has also started to play a more prominent role in the lives of children, becoming a part of their daily experience. It is now commonplace for children to interact with VAs like Siri or Alexa to open an app, answer questions, get directions, or play their favorite songs. In fact, over a third of parents surveyed said that their child under 11 years old does this regularly (Pew Research Center, 2020), and it seems likely that this proportion will continue to increase in the years to come. This VA technology provides new opportunities for young children who may struggle to use traditional keyboards due to developmental immaturity or disabilities (Gossen et al., 2013). Furthermore, as speech recognition technology in general improves, we see these types of assistants increasingly used in educational and health related products markets for children (e.g., reading support tools and language learning apps; Terzopoulos & Satratzemi, 2020).

But how well do virtual assistants (VAs) understand children’s speech? Past research has primarily focused on small samples of school-aged children spanning broad age ranges, i.e., five years and above, with Kennedy et al. (2017) including just 11 five- to six-year-olds, and Kim et al. (2022) including 28 five- to 10-year-olds in their respective analyses. While VAs generally perform better with adult speech, their performance with child speech has been steadily improving over the years. However, limitations still exist. For instance, young children experience varying degrees of success when interacting with VAs. Most studies report that 5-year-olds’ speech is transcribed correctly less than 50% of the time (e.g., Kennedy et al., 2017; Kim et al., 2022). Although one study finds a relatively high average transcription accuracy for child-led conversations at approximately 89%, devices only returned meaningful on-topic responses to children’s requests about half of the time (Lovato et al., 2019). Indeed, VA transcription accuracy is dependent on the age of the child (Kim et al., 2022). Notably, no study has examined VA transcription accuracy longitudinally within the same children as they grow older. Such an approach could provide more insight into how transcription accuracy evolves with children’s developing speech, compared to cross-sectional data on children of different ages.

Many questions remain regarding VA’s ability to handle child speech. At least to human listeners, children under five are far more difficult to understand than school age children (Hustad et al., 2020; Yu et al., 2023). Despite this, there is a notable lack of data on how VAs manage the speech of children younger than five years old. These inquiries are crucial, especially as voice technology becomes increasingly prevalent in children's daily lives and is often promoted as a tool to assist children requiring additional educational support (Hiniker et al., 2020; Kim et al., 2021).

Is it reasonable to presume that voice recognition technology can handle the speech of toddlers and preschool children if it has recently shown adequate (though not tremendous) performance with the speech of school-aged children aged five? Are VAs increasingly improving in their performance with children as children themselves become more frequent users of this technology? These are some of the questions that the literature currently cannot address.

As a first step towards addressing these gaps in the literature, we examine the performance of Siri and Alexa, two popular voice assistants readily available to children on their caregivers’ phones, on its ability to transcribe speech produced by children and adults. To see how child age affects VA’s performance, we present Siri and Alexa with recordings of the same children made when the children were two-, three-, and five-years-old. The adult speech will all be that produced by child participant’s mothers. To see how much Siri has improved over the years in its performance with children, we test an older and newer version of Siri. To gauge how Alexa and Siri’s performance compares to human listeners, we also test a different set of mothers (unfamiliar with the child voices used in the study) on their ability to transcribe these same recordings. We also test a group of undergraduate students who do not routinely interact with young children, to test the prediction that the mothers we tested might be particularly skilled at understanding young children (due to their routine experience interacting with these age groups). Our results reveal new insights regarding VA’s ability to handle child speech and lay the groundwork for future studies examining VA performance with different populations (e.g., children who are learning English as a second language or who have developmental delays that may impact their intelligibility).

Based on the limited literature, our predictions are as follows: 1) Both mothers and undergraduates will outperform all three VA systems in their ability to accurately transcribe children’s speech, 2) VA systems (especially the older Siri model) will perform far worst with the speech of two-year-old children than the speech of older children and adults (suggesting past studies have overestimated how well VA systems perform with children’s speech by testing older children), and 3) if routine exposure to children makes them easier to understand, mothers may outperform undergraduates on child (but not adult) speech. However, if child speech is inherently difficult to understand (e.g., due to its acoustic variability; Fikkert, 2010; Lee et al., 1999; MacDonald et al., 2012; Vihman, 1993), mothers and undergraduates may struggle equally with child speech – particularly when transcribing the youngest children. Testing this third prediction will provide insight into the question of how high our expectations can be for VA systems to improve through experience in their ability to understand child speech.

Method

*2.1 Participants*

Five listener types were tested in the current study: young adults (*N* = 48), mothers of young children (*N* = 48), an older version of Siri (‘Siri 2019’, OS 13.3) and a newer version of Siri (‘Siri 2024’, OS 17.3.1), and the current version of Alexa (version 2024.21). Undergraduates (*M*age = 19.25 years; 36 female; 12 male) from the Greater Toronto Area were tested and received course credit or $10 compensation for participating. The mothers (*M*age = 37.67 years; 48 female) were invited to participate while their child was completing another study in-person in the lab. All participants learned English before the age of six in North America and English was their dominant language. All the participants had normal hearing and normal or corrected-to-normal vision.

*2.2 Stimuli*

The speech stimulus used in this study were longitudinally recorded utterances of 32 words by the same twelve children at three age points:two,three, and five years – and by 12 mothers of similarly aged children who were all native speakers of English with no perceptible non-local accent. The children were monolingual and learning English as their only language. The mothers were also monolingual and had learned English before the age of six in North America. The 32 words are typically known by 30-month-olds (Frank, Braginsky, Yurovsky, & Marchman, 2016). Examples include cow, plane, monkey, ball, and horse (See appendix for full word list). Recordings were made in a double-walled, sound attenuated Industrial Acoustics Company (IAC) booth using high-quality recording equipment and stored in WAV Audio File Format (48kHz; normalized to 69.5 dB). The initial set of recordings completed when the children were two years of age, were the same as those used by Cooper et al., (2018). These same children were then followed longitudinally at three and five years. For the children, the words were elicited in a child-friendly experimenter-controlled video game, which allowed for a spontaneous elicitation of the word without repetition.

*2.3 Transcription study design*

In the adult transcription task, each listener heard all 32 words produced by four different talkers (i.e., three different children and one of the twelve adult mothers). Eight productions were heard from each talker, and no word was heard twice by a single participant. The productions were presented in a randomized order and all child recordings were heard across participants.

For the ASR transcription task, a 9.7-in. tablet computer (Apple iPad, 2018) with automatic speech recognition (ASR) technology, with the two versions of Siri (OS 13.3 and OS 17.3.1) and the Amazon app with Alexa (2024.21), were used for the ASR transcription task. The full set of child and adult mother recordings were presented through speakers while Siri was activated through the Notes application, and Alexa was activated through the Amazon app, by an experimenter in the lab.

*2.4 Procedure*

Each participant completed the self-paced transcription task one-on-one, in person, in a sound attenuated booth in our lab space. The entire testing session took about five minutes for each participant. The task was created and run using the online experiment builder Gorilla (www.gorilla.sc) and presented using a 13-in. laptop (MacBook Air, 2017). Auditory stimuli were presented over Sennheiser HD PRO headphones at a constant, comfortable listening level and results were stored in the “Firebase” cloud database. The participants were presented with each word over the headphones and were asked to transcribe each production by typing into an onscreen text box. Once a transcription was entered, participants were able to advance to the next trial. No feedback on their transcription accuracy was provided. For Siri and Alexa, the procedures were replicated so that each version of Siri (OS 13.3 and OS 17.3.1) and the version of Alexa (2024.21) was tested on all the conditions completed by the human participants. The ASR transcription task took place in an IAC sound-attenuated booth while an experimenter used either Siri or Alexa.

Siri was activated within Apple’s Notes application on the tablet computer while the recordings were played over a speaker. Siri was activated prior to each word and deactivated following each transcription into the Notes app. Siri’s dictation history was cleared after each of the full sets of 32-word to ensure that previous transcriptions did not influence or bias the results of subsequent presentations, allowing for an accurate and independent assessment of Siri’s performance under each condition. Alexa was used on the Amazon Alexa app on the iPad to listen to the words, and the transcription was found in the voice active history on the app. The history activity was deleted after each word was presented. Instances when no transcription followed the presentation of a word was marked by an ‘x’ or ‘.’. Siri and Alexa’s transcription errors were manually noted. Both human participants and VA transcriptions were coded as ‘correct’ if they matched the target word or ‘incorrect’ if the transcription did not match the target word (e.g., a different word, a statement, no transcription, etc.). **To ensure comparability, each VA was tested 48 times, just like the adult listeners, making the data directly comparable across both groups.**

Results

Transcription accuracy of children and adult talkers by human listeners (undergraduate and mother), Alexa, and the two versions of Siri (Siri 2019 and Siri 2024) were compared across talker age (see Figure 1). We employed a generalized logistic mixed-effects regression model to our data using the *glmer* function in the *lme4* package Version 1.1-21 (Bates et al., 2015) in R. The model included the binary outcome variable (1 = correct response; 0 = incorrect response), and the independent categorical variables of listener type (Siri 2019, Siri 2024, Alexa, mothers, and undergraduates) and talker age (two, three, five years, and adult). The model also included a random intercept for item and participant, and a random by-participant slope for talker age. Since we expected listeners with greater experience with child speech to perform better compared to those with less, listener type was forward-difference coded to allow for the adjacent comparisons: 1) mothers versus undergraduates, 2) undergraduates versus Alexa, 3) Alexa versus Siri 2024, and 4) Siri 2024 and Siri 2019. Talker age was coded using Helmert contrasts to compare transcription accuracy between: 1) adults versus children (5-, 3-, and 2-year-olds), 2) 5-year-olds versus 3- and 2-year-olds, and 3) 3- versus 2-year-olds. The formula for the model is as follows:  
 Correct ~ Listener + Age + (1 | Item) + (1 + Age | Subject)

As predicted, we found that transcriptions were less accurate for younger children than for older children and adults, and that humans were more accurate at transcribing both children and adults than VA’s. The model indicated significant main effects for both listener type and talker age (***p* < .001**), with improving word recognition across all listeners as speaker age increased, such that the highest transcription accuracy was observed for adult speech (*M* = .94; *SD* = .25); that is, all listeners transcribed speech by adults better than speech by children (Estimate: 2.6296, *SE* = 0.260, *z* = 10.106, *p* < .001, *OR* = 13.869), speech by 5-year-olds (*M =* .86; SD = .35) was transcribed better than speech by 3 (*M* = .70; *SD* = .46) and 2-year-olds (M = .57; SD = .49; Estimate: 1.7605, *SE* = 0.126, *z* = 13.968, *p* < .001, *OR* = 5.815), and speech by 3-year-olds were transcribed better than speech by 2-year-olds (Estimate: 0.97982, *SE* = 0.127, *z* = 7.690, *p* < .001, *OR* = 2.664). When collapsed across all talker ages, mothers achieved the highest transcription accuracy (*M =* .900*; SD* = .300), however their performance did not significantly differ from that of undergraduates (*M* = .873; *SD* = .333; Estimate: 0.5419, *SE* = 0.327, *z* = 1.659, *p* = .0972, *OR* = 1.719). Undergraduates transcribed speech more accurately compared to Alexa (*M* = .676; *SD* = .468; Estimate: 1.9028, *SE* = 0.202, *z* = 9.402, *p* < .001, *OR* = 6.704). While transcription accuracy did not differ between Alexa and Siri 2024 (*M* = .689; *SD* = .463; Estimate: -0.0900, *SE* = 0.150, *z* = -0.600, *p* = .5487, *OR* = 0.914), Siri 2024 did outperform Siri 2019 (*M* = .585; *SD* = .493; Estimate: 0.6265, *SE* = 0.150, *z* = 4.175, *p* < .001, *OR* = 1.871). There were no significant interactions between talker age and listener type; however, results suggested that human listeners (e.g., mothers, undergraduates) generally outperformed VAs (e.g., Alexa, Siri 2024, Siri 2019). Therefore, to further investigate this difference in performance, we constructed another logistic mixed-effects regression model that followed the same structure as the model above, with listeners grouped by listener type: human vs. VAs (with humans as the reference level). In line with our predictions, humans (*M* = .886; *SD =* .317) outperformed all VAs (*M* = .651; *SD =* .477) in transcribing child and adult speech. We again observed no significant interactions between talker age and listener type, indicating that human listeners generally outperform VAs in transcribing speech across all talker age points.

While the broader model did not reveal significant interactions between age and listener type, we had predicted that listener familiarity with the youngest children’s speech might offer a performance advantage (i.e., mothers might transcribe two-year-olds speech more accurately than undergraduates due to their frequent interactions with young children). To examine whether listener experience confers an advantage with the youngest speakers, we constructed a logistic regression model specifically focusing on two-year-olds’ speech. Listener type was simple coded (with mothers as the reference level). A random intercept was included for item and participant. The model revealed that mothers outperformed all other listener types with the youngest age group. Mothers, who achieved the highest average transcription accuracy (*M* = .77, *SD* = .42), significantly outperformed undergraduates (*M* = .69, *SD* = .47; Estimate: -0.546, SE = 0.276, z = -1.975, p = .0483, *OR* = 0.579). Mothers also outperformed Alexa (*M* = .65, *SD* = .48; Estimate: -1.9398, SE = 0.277, z = -7.014, p < .001, *OR* = 0.144), Siri 2024 (*M* = .43, *SD* = .50; Estimate: -1.8937, SE = 0.276, z = -6.860, p < .001, *OR* = 0.151), and Siri 2019 (*M* = .27, *SD* = .45; Estimate: -2.7274, SE = 0.285, z = -9.588, p < .001, *OR* = 0.065). To visualize how VA’s and mothers performed on specific items in our study, we examined the transcription accuracy of words produced by two-year-old children by mothers and VA systems (see Figure 2). To point to potential differences in the types of errors humans and VA’s make with child speech, we compiled a table reviewing the average percentage of errors and top five error-prone words for human (both young adults and mothers combined) and VAs (Siri 2019, Siri 2024, and Alexa combined) based on two-year-olds’ speech (see supplementary materials).

A graph of different colored lines

Description automatically generated

Fig. 1. Average transcription accuracy of children and adult talkers by listener type (error bars indicate by-participant standard error)

A graph of different colored lines

Description automatically generated with medium confidence

Fig. 2. A plot by item of transcription accuracy on words produced by 2-year-olds between mothers, Alexa, and Siri 2024 (error bars indicate by-participant standard error).

1. Discussion

VA technology is a huge part of our everyday lives and is becoming increasingly more integrated into the lives of children (Kim et al., 2022). Nevertheless, most research in this area has focused on how VAs cope with the speech of monolingual adults. The little work that has been done with children has involved typically developing children aged five and above (e.g., Kennedy et al., 2017; Kim et al., 2022; Lovato et al., 2019). And even then, these studies have often collapsed data collected from children spanning a broad age range and/or used recordings of only a small number of child talkers. This present study is the first to compare different types of adult and VA listeners’ comprehension of a large number of children under five years of age.

Our results underscore the need for further research into how voice technology can adapt to the challenges posed by children's speech, especially as it becomes increasingly integrated into educational contexts and daily life. The new version of Siri did indeed outperform the older version of Siri but still showed significantly lower accuracy with children’s speech compared to adult speech. Specifically, both Siri and Alexa's recognition accuracy decreased with younger child speakers, with hit rates for two-year-olds being the lowest. This shows that updated VA technology is moving in the right direction, but further refinement is needed to improve accuracy with young children. Human listeners outperformed Siri and Alexa for both child and adult speech, with a more significant performance gap observed for child speech. This Siri/Alexa-human performance gap was most pronounced with the two-year-old speakers which provides evidence for the claim that humans are better than ASR systems at understanding children (e.g., Gerosa et al., 2009), especially at the younger ages. However, it is important to note that even human listeners faced challenges when transcribing the speech from the youngest children.

Why is children’s speech so difficult to understand? These difficulties may be linked to the variability and unpredictability inherent in child speech (e.g., Fikkert, 2010; Lee et al., 1999; MacDonald et al., 2012). For example, young children often produce speech that deviates from adult-like phonological norms, such as omitting phonemes or syllables, or simplifying complex sound combinations, variable articulation, and idiosyncratic substitutions that deviate from adult-like norms (Vihman, 1993). Perhaps this inherent variability, along with factors such as immature control over prosody, timing, and pitch (e.g., Foulkes et al., 2005), contributes to the difficulties faced by both human listeners and voice recognition technology. Beckman et al., (2018) further argues that children’s vocal tracts are smaller than that of an adult and rapidly change in size and shape throughout development which contributes to between-talker variability that dwarfs the difference between adult speakers. This is coupled with the fact that child vocabularies and phonological proficiencies are developing which leads to increased within-talker variability.

One could presume that VAs may struggle with children’s speech because they are predominantly trained on datasets of adult voices. The fact that we found that mothers outperformed undergraduates, albeit with only a slight advantage, in understanding the speech of two-year-olds is in alignment with this explanation. Thus, future work could seek to optimize VA performance with children’s speech by increasing VA technology’s exposure to child speech. But given that even mothers with young children performed far better with adult speech than two-year-old speech, it may be an unrealistic goal to ever expect VA technology to perform equally well with child and adult speech.

Another important avenue for future work will be to explore the generalizability of these findings by testing VAs on multi-word utterances with different levels of background noise, since children’s interactions with VA’s are often not isolated words recorded in sound booths. By doing so, we may gain a better measure of how these technologies handle real-word input from children. Another important future direction will be to test these systems on the speech of children who speak English as a second language, or who have a developmental delays. By better understanding how these factors impact VA technology’s success with children’s speech, we will better understand how to develop more effective voice technology solutions that cater to the needs of all children.

To summarize, this study highlights the current limitations of VA technology in understanding young children’s speech, and underscores the need for additional research in this area. Our findings are consistent with the notion that more training with young children’s speech might help develop more effective voice technologies that cater to the needs of children, but they also suggest that VA technology performance with children’s voices may never reach the accuracy we see with adult voices.

Supplementary Materials

See supplementary material at [URL will be inserted by AIP] forthe following table: All 32 Words, Listed in Order from Least Accurate to Most Accurate (for Two-Year-Old Speech, as Judged by Joint Average of Undergrads and Moms). This Table presents average transcription accuracy from humans (averaged across mothers and undergraduates) versus voice assistant (VA) technology (averaged across Alexa and the more recent version of Siri). Note, the top five incorrect errors are given except for when there were less than five types of errors.

Acknowledgements

This research was supported by grants awarded to E.K.J. from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada.

Author Declarations

*Conflict of Interest*

The authors have no conflicts of interest to report.

*Ethics Approval*

This study was approved by the University of Toronto Research Ethics Board. Informed consent was obtained for all participants through their caregivers.

Data Availability

The data supporting this study's findings are available from the corresponding author upon reasonable request.

References

Basak, S., Agrawal, H., Jena, S., Gite, S., Bachute, M., Pradhan, B., & Assiri, M. (2023).

Challenges and limitations in speech recognition technology: A critical review of speech signal processing algorithms, tools and systems. Computer Modeling in Engineering & Sciences, 135(2), 1053–1089. <https://doi.org/10.32604/cmes.2022.021755>

Beckman, M. E., Plummer, A. R., Munson, B., & Reidy, P. F. (2017). Methods for eliciting, annotating, and analyzing databases for child speech development. *Computer speech & language*, *45*, 278–299. <https://doi.org/10.1016/j.csl.2017.02.010>

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Ed.), *Laboratory Phonology 10* (pp. 227-260). Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110224917.3.227>

Foulkes, P., Docherty, G. J., & Watt, D. (2005). Phonological variation in child-directed speech. Language, 81(4), 799–823. https://doi.org/10.1353/lan.2005.0237

Gerosa, M., Giuliani, D., Narayanan, S., & Potamianos, A. (2009). A review of ASR technologies for children’s speech. In *Proceedings of the ISCA Archive Computer Interaction (WOCCI 2009)* (pp. xx-xx). Cambridge, MA, USA, November 5, 2009.

Gossen, T., Kotzyba, M., Stober, S., & Nürnberger, A. (2013). Voice-controlled search user interfaces for young users. In 7th Annual Symposium on Human-Computer Interaction and Information Retrieval (pp. 2–5)

Hiniker, A., Druga, S., & Szewczyk, P. (2020). The role of voice technology in supporting children with developmental disabilities. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW3), 1–24.

Hustad, K. C., Schueler, B., Schultz, L., & DuHadway, C. (2012). Intelligibility of 4-year-old children with and without cerebral palsy. Journal of Speech, Language, and Hearing Research, 55(4), 1177–1189.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E.& Belpaeme, T. (2017). Child speech recognition in human-robot interaction: Evaluations and recommendations. HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, 82 90.

Kim, M. K., Druga, S., Esmaeili, S., Woodward, J., Shaw, A., Jain, A., Langham, J., Hollingshead, K., Lovato, S. B., Beneteau, E., Ruiz, J., Anthony, L., & Hiniker, A. (2022). Examining voice assistants in the context of children’s speech. University of Washington

Kim, S.-E., Chernyak, B. R., Seleznova, O., Keshet, J., Goldrick, M., & Bradlow, A. R. (2024). Automatic recognition of second language speech-in-noise. JASA Express Letters, 4(2), 025204.

Kuhn, K., Kersken, V., Reuter, B., Egger, N., & Zimmermann, G. (2024). Measuring the accuracy of automatic speech recognition solutions. ACM Transactions on Accessible Computing, 16(4), Article 25.

Lee, S., Potamianos, A., and Narayanan, S. (1999). “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” *Journal of the Acoustical Society of America, 105,* 1455–1468

Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019). Hey Google, do unicorns exist? Conversational agents as a path to answers to children’s questions. In Proceedings of the 18th ACM International Conference on Interaction Design and Children (pp. 301–313). ACM.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., and Munhall, K. G. (2012). “Children’s development of self-regulation in speech production,” *Current Biology, 22*(2), 113–117

Pew Research Center. (2020, July). Parenting children in the age of screens. Pew Research Center.

Proksch, S.-O., Wratil, C., & Wäckerle, J. (2019). Testing the validity of automatic speech recognition for political text analysis. Political Analysis, 27(3), 339–359. <https://doi.org/10.1017/pan.2018.62>

Terzopoulos, G., & Satratzemi, M. (2020). Voice assistants and smart speakers in everyday life and in education. Informatics in Education, 19(3), 473–490. <https://doi.org/10.15388/infedu.2020.21>

Vihman, M. (1993). Variable Paths to Early Word Production. *Journal of Phonetics, 21,* 61-82.

Yu, M. E., Cooper, A., & Johnson, E. K. (2023). Who speaks "kid?" How experience with children does (and does not) shape the intelligibility of child speech. *Journal of experimental psychology. Human perception and performance*, *49*(4), 441–450. <https://doi.org/10.1037/xhp0001088>