Universitat
Pompeu Fabra
Barcelona

Treball de fi de màster: *Recerca*

Development of vocal repertoires in non- and minimally verbal autism spectrum disorder

Nom i Cognoms: *Kelly Lauren Sheets*

Màster: *Lingüística Teòrica i Aplicada*

Edició: 2021-2022

Director o directora: *Dr. Wolfram Hinzen*

Any de Defensa*:* 2022

Col.lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Wolfram Hinzen, without whom I would not have the confidence to discuss or write about such topics in any coherent capacity. I am sincerely thankful for all the time you have invested in me over the last few months, and most of all, for being open to supervising a fashion major from Texas.

I'd also like to extend the same sentiments to my co-supervisor, Dr. Dominika Slušná. I am humbled that you have trusted me to work with a population that is near and dear to your heart. Your passion for research is contagious, and I am overjoyed at the opportunity to have learned from you. I hope this will not be the last time.

I'd like to thank Dr. Erick Canales-Rodríguez, for the patience and willingness to send the necessary code and databases to complete this project. Your slides on fractional anisotropy and TractSeg were passed down through a generation of supervisees.

Thank you to the entire Grammar and Cognition group for allowing me to work alongside you on campus, perhaps against your will at times. I have been moved by your eagerness to help in my work. You have unknowingly contributed to this thesis in many ways.

I'd like to also thank my friends at home, but especially those I've made within the four walls of UPF. Thank you for being my biggest cheerleaders and for shared laughter, tears and memories. I will feel a deep sense of loss when we part ways towards our next chapters, but will cherish what I know will be lifelong friendships.

Most of all, I would like to send my appreciation to my loving family including my mom, dad, twin brother and two dogs. For my parents especially, from the bottom of my heart I want to thank you for never making me feel like there was ever a wrong time or right way to live my dreams. I'm at a loss for words of how to express my gratitude for your endless sacrifices, encouragement and support. Thank you for showing patience even when I was so stressed, I only wanted to FaceTime with the dogs.

Abstract

The autism spectrum comprises a substantial 30% of individuals said to be non- or minimally verbal. While these produce no phrase-speech by definition, they often do vocalize. However, their exact vocal repertoires remain unknown. The present study aimed to advance on this front by developing a new annotation scheme for nine such individuals, adapting categories used to profile pre-verbal neurotypical children. In a second step, vocalization variables were related to microstructural metrics (fractional anisotropy, FA, and myelin water fraction, MWF) of a crucial white-matter speech-related tract, the arcuate fasciculus (AF). Results revealed that at a group level, participants produced significantly more syllabic, or speechlike, than non-syllabic vocalizations. The proportion of words marginally exceeded that of word approximations. Furthermore, the vowel and consonant repertoire resembled that of neurotypical infants. There were no significant correlations between vocalization and either FA or MWF in the AF. These new findings refute the expectation that vocalizations in individuals with NMVA are predominantly non-speech-like and substantiate the need to look for the neural basis of NMVA also in non-speech-related language territory in the brain.

Keywords: non- and minimally verbal autism, vocal repertoire, arcuate fasciculus, MWF, FA

Table of Contents

Acknowledgments……………………………………………………………………………1

Abstract…………………………………………………………………………………..2

1. BACKGROUND…………………………………………………………………….5

    1.1. The linguistic phenotype of non-verbal and minimally verbal autism spectrum disorder (NMVA)………………………………………………………………….5

    1.2. The trajectory towards first words…………………………………………………7

    1.3. The pre-linguistic stage and eventual language outcomes in ASD and NMVA….....10

    1.4. Study aims…………………………………………………………………...12

    1.5. Research questions…………………………………………………………..12

2. STUDY 1: THE VOCAL REPERTOIRE OF 9 YOUTHS DIAGNOSED WITH NMVA……………………………………………………………………...…13

    2.1. Methods…………………………………………………………………...13

        2.1.1. Participants………………………………………………………………13

        2.1.2. Annotation scheme design……………………………………………….14

        2.1.3. Coding criteria……………………………………………………………18

        2.1.4. Materials…………………………………………………………………19

        2.1.5. Analysis…………………………………………………………………..20

            2.1.5.1. Speechlike complexity across the group…………………………….20

            2.1.5.2. Vocal repertoire by subject……………………………………........21

        2.2. Results…………………………………………………………………...…22

            2.2.1. Speechlike vocalizations………………………………………………...22

            2.2.2. Vocal repertoire………………………………………………………..24

3. STUDY 2: RELATION OF VOCAL REPERTOIRES WITH WHITE MATTER INTEGRITY OF THE AF……………………………………………………………26

3.1. Methods……………………………………………………………………….26
3.1.1. Participants……………………………………………………….26
3.1.2. Diffusion MRI: data acquisition, preprocessing and estimation………26
3.1.3. Multi-Echo T2 Relaxometry: data acquisition, preprocessing and estimation……………………………………………………………………27
3.1.4. Tract segmentation………………………………………………28
3.1.5. Analysis of FA and MWF in the AF……………………………..28
3.2. Results of FA and MWF in the AF…………………………………..……....30
4. DISCUSSION…………………………………………………………………..31
4.1. Limitations…………………………………………………………….,.35
5. CONCLUSION………………………………………………………………....35
6. REFERENCES…………………………………………………………………36
7. APPENDIX A…………………………………………………………………..53

1. Background

1.1. The linguistic phenotype of non-verbal and minimally verbal autism spectrum disorder (NMVA)

Humans have an innate ability to perceive and produce speech as a result of complex neural and biological endowments, not fully replicated in non-human primates (Dehaene-Lambertz et al., 2006; Oller et al., 2019; Peter 2013; Kent 2022;); yet there are some individuals with rare and common neurodevelopmental disorders in which speech and language fail to develop. One such phenotype with highly heterogeneous linguistic behavioral expression is autism spectrum disorder (ASD), composed of a subset estimated ranging from 25%-50% (Norrelgen et al., 2015; Hus et al., 2007; Pickett et al., 2009) that will remain non- or minimally-verbal past school age (Norrelgen et al., 2015). While definitions and criteria for what constitutes non- or minimally verbal are mixed in literature and research (for a review of definitions of non- and minimally-verbal ASD see Koegel et al., 2020), those diagnosed with non- or minimally verbal autism (herein referred to as NMVA) produce speech only amounting to a few words or fixed phrases by school age (Yoder and Stone 2006). Production deficits are not compensated for through other means such as gesture or prosthesis (La Valle et a., 2021; Gernsbacher 2004). Receptive language has also been found to be highly variable and impaired across the entire spectrum (Rapin et al., 2009) and in NMVA, may range from comprehension of single words to a few fixed phrases. Plesa Skwerer and colleagues (2016) found over half of minimally verbal ASD children tested on the Peabody Picture Vocabulary Test -4 (PPVT-4), a picture-word matching test, were at floor or unable to reach baseline and ranged from 1.4 to 4 standard deviations below the mean of the average neurotypical score. In a study comparing linguistic and cognitive profiles of youngsters and adults with NMVA, Slušná and colleagues (2021) sample fell 2 standard deviations below the neurotypical average on the PPVT-3. Moreover, receptive language ability not only correlated with, but predicted, expressive language ability in both demographics. This expressive and receptive profile was found to be similar across both age groups, suggesting that the linguistic profile of NMVA is unlikely to change over the course of time (Slušná et al., 2021).

Physical impairment cannot fully elucidate the linguistic confounds in NMVA, as gestures and/or prosthetic devices that might aid in signing or written communication do not appear to assist such cases. In the case of non-verbal ASD, the auditory network, perception and processing remain largely intact – even showing leftward lateralization as with neurotypical infants, although lowered Blood Oxygen Level Dependent responses to speech stimuli have been detected (Slušná 2021; Dehaene-Lambertz et al., 2002; Linke et al., 2021). Responses to speech-based stimuli ASD were documented to be the same as with typically developing controls (Schwartz et al., 2018), indicating that no deficits have been found yet that can explain such severely impacted language by means of a deconstructed phonological loop. Recent research has expounded upon a small subgroup within NMVA suffering from oro-motor deficits including childhood apraxia of speech (CAS) and dysarthria (Chenausky et al., 2019; Tierney et al., 2015). Dysarthria, a neuromuscular disorder affecting muscle tone in the jaw and subsequently the coordination and movement needed to verbalize, does not serve as a sole explanatory factor for even expressive language problems, as the little language that those with NMVA produce also lacks substantial referentiality (Hinzen et al., 2019). The same can be said for CAS, a developmental speech-sound disorder, which impedes articulatory precision and control (ASHA 2007; Chenausky et al., 2017). Both diagnoses fall short of illuminating the global linguistic impairment in NMVA: vocabulary acquisition, arbitrary semantic usage, the inability to communicate via signed or written modalities and hindered receptive language. Nevertheless, it is difficult to exclude any rationale for atypical linguistic outcomes in NMVA, as the disorder is characterized by an extremely high rate of comorbidities. Levy et al. (2010) collected data on 2,568 diagnosed ASD individuals and observed that 82.7% suffered from other comorbidities including unrelated developmental diagnosis, psychiatric diagnosis, neurologic diagnosis and a contributing genetic diagnosis.

The primary aim of this thesis is to characterize developmental outcomes in speech production in NMVA through a detailed phonological analysis. Building a phonological repertoire of NMVA presents a formidable challenge, as these adults, teenagers and children whose vocalizations most resemble pre-verbal infants, still remain functionally different in their speech production. Nonetheless, what is known about the neurotypical infant's pathway towards the heralded first words can inform the foundations of an annotation scheme developed to describe the vocalizations of NMVA.

1.2. The trajectory towards first words

While the pathway to speech is invisible to the naked eye, it is a long and intensive process that begins after fertilization. During gestation, a fetus is harbored in an auditorily and acoustically diverse space that provides input from external noise and voicing (Peter 2013). Auditory perception develops around the second trimester, as detected by fetal movement and brain responses to auditory stimuli (Hepper and Shahidullah 1994; Draganova et al., 2007). Lower frequencies like vowels and voiced consonants, are first more permeable to the fetus in comparison with higher frequency voiceless consonants (Gerhardt and Abrams 2000). The perisylvian region responsible for crucial language processes, including the inferior frontal lobe and temporal lobe, are discernable at 6 months gestational age (Mahmoudzadeh et al., 2013). At birth, the infant's respiratory system will make its first ingressive-egressive sequence in order to let out their first advancement in linguistic development that also serves as an important sign of life, immediately after birth – the cry, which is already shaped by exposure to prosodic features of the ambient language (Mampe et al., 2009).

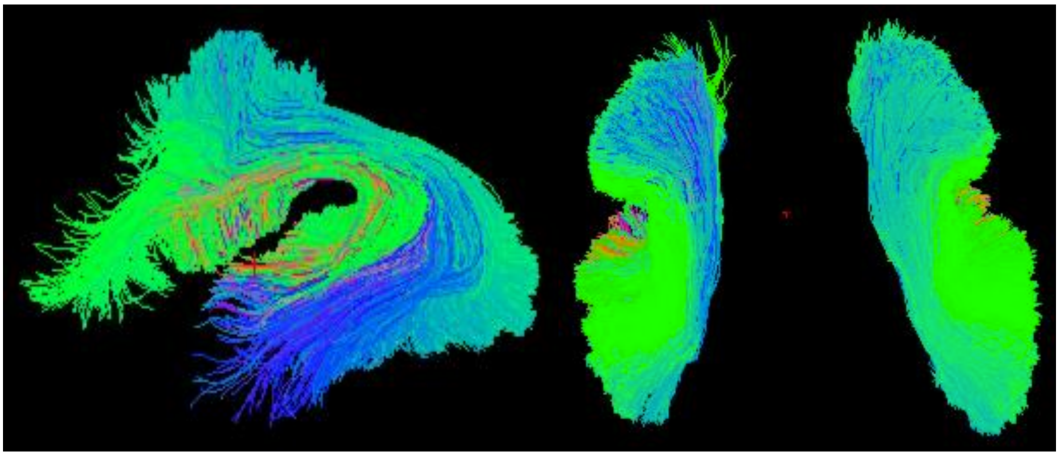
Around 2 to 4 months of life protophones emerge, i.e. pre-speech vocalizations (Oller et al., 2019), in corresponding order to the aforementioned sound frequencies that permeated the in-utero environment. These vocalizations will steadily increase in articulatory and acoustic complexity, or become more *speechlike* (Oller 2000), coinciding with social, vocal tract and neurological maturation. Phonated protophones will emerge first, as the teeth and other articulatory structures, which are needed to form articulated and syllabic vocalizations, are absent or not yet in their adult-like position (Peter 2013). Infants will begin with vowel-like, or vocalic, sounds, and later consonant-like, or closant sounds (McLeod and Bleile 2003). The quasivowel, a foundational phonated protophone crucial for later canonical babbling and speech, emerges soon after birth and has been found to be longer in duration in infants with severe disorders (Lynch et al., 1995). From 6-8 months, myelination of the left hemisphere frontal and temporal lobes, which are key language regions, rapidly occurs and will parallel the timing of language acquisition (Pujol et al., 2006). The variety of an infant's vocal repertoire at this stage has also been found to predict their productive vocabulary at 1 year (Werwach et al., 2021). As vocalizations become more tactile and exploratory, infants are able to realize the motor-speech representations of their target language (Schoen et al., 2011), which culminate in the long-awaited period of canonical babbling, a syllabic consonant-vowel or vowel-consonant vocalization with a quick transition and timing that resembles mature speech, usually emerging around 7 months of age. Anomalous canonical babbling has been found to serve as a potential early identification of normally late detected developmental disorders (Lang et al., 2019). Likewise, the late onset of canonical babbling and the size of a child's phonetic repertoire have been found to be clinically significant (Eilers and Oller 1994; Davis & Velleman 2000).

By the end of an infant's first year of life, neurotypical infants will have acquired the phonetic repertoire of their native language (Dehaene-Lambertz et al. 2006; Werker and Tees

1984). The first words will appear between 12 months to 18 months, coinciding with white matter maturation of multiple language-related areas including the superior temporal gyrus, as measured by Fractional Anisotropy (FA) and Myelin Water Fraction (MWF) (Su et al., 2008; Deoni et al., 2015). FA is a blanket measure of microstructural integrity of white matter that has been shown to parallel myelination in the first year of life and increase continuously through young adulthood (Gilmore et al., 2018). MWF, as with FA, is a microstructural marker more sensitized to myelin alone, and in a cohort of infants and young children ages 2 months to 48 months, demonstrated a significant relationship to receptive and expressive language ability over other cognitive measures (O'Muircheartaigh et al., 2014).

In a dual stream language model (Hickock and Poeppel 2004) one such dorsal white matter tract, the arcuate fasciculus (AF), matures in microstructure enabling advancement in phonological awareness, vocabulary knowledge and the fine-tuning of language exposure via conversational turn – all of which lead to increased language skills seen at 4 to 5 years old (Zuk et al., 2021; Saygin et al., 2013; Skeide et al., 2016; Romeo et al., 2018). From kindergarten until 8 years of old, age-related microstructural changes to the AF can predict receptive and expressive language. Beyond childhood, the AF has been implicated in a compendium of speech functions including speech repetition, improvised speech, and sound perception to articulatory movement (Catani et al., 2005; Saur et al., 2008).

Figure 1. Two Diffusion Tensor Imaging (DTI) depictions of the AF: a lateral view of the left AF (left), and an axial view of the bilateral AF (right).



Note. Anatomically, the AF connects the perisylvian areas of the temporal, parietal and frontal lobes (Catani and Thiebaut de Schotten 2008). DTI images, as with above, are color coded by directionality of the fiber bundles: posterior- anterior in green; superior-inferior in blue; left-right in red.

1.3. The pre-linguistic stage and eventual language outcomes in ASD and NMVA

Diagnosis of any form of ASD often occurs around four years old (Maenner et al., 2020), as there are no associative physical characteristics that serve as an early warning sign. The first warning signs are often exhibited by atypical vocalization patterns observed by parents towards the end of the first year of life, as well as neurofunctional deviances (Marschik et al., 2017). An infant vocalization profile on the NMVA phenotype remains elusive; however, existing work on ASD can inform what might be expected in the non- or minimally verbal population.

In a home-based longitudinal study, DeVeney and colleagues (2021) revealed no significant findings for frequency of vocalizations, frequency of speechlike vocalizations, consonant inventory, non-speechlike vocalizations or percent of canonical syllables between infants at high and low risk for ASD at 6 months of age; however, at 12 months the low risk surpassed the high-risk group in vocalization rate, but did not notably differ on other measures. This same progression was confirmed by an additional longitudinal study also observing disparities at 12 months between risk groups, not at 6 months, which correlated with diagnosis, receptive and expressive language outcomes and autism severity (Plate et al., 2021). Infants later diagnosed with ASD were late and less frequent canonical babblers (Cychosz et al., 2021) at 9 to 12 months in comparison with typically developing infants (Patten et al., 2014). Babbling in typically developing infants and those with ASD can predict expressive language ability (McDaniel et al., 2018), vocabulary size at 18, 24 and 30 months as well as receptive vocabulary at 12 and 18 months (Oller 199; Sheinkopf et al., 2000; Mitchell et al., 2006).

In one study of minimally verbal children ranging from 2 to 4 years old, consonant inventory and intentional communication were two predictors of expressive language growth; predictors of receptive vocabulary growth also included intentional communication, but not consonant inventory (Yoder et al., 2015). In a slightly older cohort of minimally verbal children, from 2 to 5 years years old, consonant inventory also correlated with later expressive language levels, but not receptive, as well as phonetic repertoire and alphabet score (Saul and Norbury 2020). Building on Broce's and colleagues (2015) work on the AF, Chenausky et al., (2017) assessed these two tracts in relation to speech and fluency in ten minimally verbal children. Speech fluency fractional anisotropy (FA) values correlated with the left AF predicted initial consonants correct in imitated speech. In a volume-based metric, a measure purported to reveal macrostructural properties white matter, 5 non-verbal children displayed right lateralization of the AF as opposed to left lateralization in typically developing controls (Wan et al., 2012).

Today research within the NMVA population still remains at the precipice and a primary research need. Recruitment remains a particular hurdle, which results in small sample sizes as seen in Chenausky and colleagues (2017) and Wan and colleagues (2012). Standardized tests have posed a difficulty to administer and adapt to NMVA (Chenausky et al., 2021), therefore, it can be assumed that further difficulties may arise even in a semi-structured evaluation such as the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012). The ADOS is a play-based session in which an examiner assesses the social, communicative and embodiment skills of those suspected of having NMVA or broader ASD diagnosis. Even in a semi-structured assessment such as the ADOS, the known social and pragmatic deficits that underscore the entire spectrum of this disorder (Tager-Flusberg et al., 2009), NMVA individuals may relate to both the examiner and environment in a highly atypical manner that inherently make any observation and/or classification challenging.

1.4. Study aims

This study aims to shed more light on the vocalization profile of children and adolescents with NMVA. While research has waded into phonetic and lexical aspects of this phenotype as reviewed above (Chenausky et al., 2019; Saul and Norbury 2020), the variation and rate of vocalizations of different types remain unknown. At both the phonetic and lexical level, there is no current research to my knowledge about the types and tokens of the words and word approximations NMVA youths spontaneously produce, or what vowels and consonants are present in them. Following compelling evidence regarding the involvement of the AF in typical speech development in children and correlations with later linguistic outcomes (Broce et al., 2015; Zuk et al., 2021) along with emerging research on the AF in minimally verbal autism (Chenausky et al., 2017), this study also aimed to elucidate vocal variation within NMVA in relation to the AF's microstructural markers using FA and WMF.

1.5. Research questions

With these aims in mind, the following research questions were formed. The first four questions will be addressed through annotating and compiling a vocal repertoire of NMVA; the last question will be addressed by relating measures of the vocal repertoire with FA and MWF values of the left and right AF.

1. What are the proportions of speechlike to non-speechlike vocalizations in NMVA?

2. Do the more pre-speech associated protophones, quasivowels and canonical syllables, correlate with word approximations and words?
3. How many types and tokens of word approximations and words are produced by each subject and what are the vowels and consonants present in them?
4. How do measures of the vocal repertoire of NMVA map onto measures of white matter integrity in the AF?

2. Study 1: The vocal repertoire of 9 youths diagnosed with NMVA

2.1. Methods

2.1.1. Participants

The subjects were 9 NMVA youths, all of whom had previously participated in a study by Slušná (2021) and recruited from special schools and daily centers in Barcelona, Spain. At the time of initial data collection, inclusion criteria in Slušná (2021) included: (a) a previous parent or center reported ASD diagnosis; (b) absence of functional speech, defined as no words, single words or simple fixed phrases as determined by parental, school or center reports and in line with professional ASD classification; (c) no compensation for lack of speech with written or sign language modalities within the participant's home or some other context.

After inclusion criteria were met, all subjects' ASD diagnosis was re-assessed using Module 1 Preverbal/Single Words of The Autism Diagnostic Observation Schedule 2nd Edition (ADOS-2; Lord et al., 2012) for those 30+ months of age, or The Adapted ADOS (A-ADOS; Bal et al., 2020) for those 13+ years old, as well as The Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003). Use of both the ADOS-2 or A-ADOS together with the ADI-R have been found to achieve a highly accurate diagnosis in research settings (Lebersfeld et al., 2021). The ADI-R is a primary caregiver-directed interview that separates ASD from symptoms of other pervasive developmental disorders; individuals not found to produce flexible 3-word phrases are described as minimally verbal. The expressive speech ability of all subjects was based on ADOS Item A1 which takes into account spontaneous speech with communicative intent that is non-echolalic in nature. A binary classification was made deeming speech production as few-to-no-words (FNW) or some-words (SW). Subjects producing few to no words had less than five types of words or word approximations during the ADOS session, and those with some words produced more than five types of words or simple phrases. Subject #7 was fluently echolalic and Subject #6 also had a comorbid genetic disorder (22q11.2 deletion syndrome). Table 1 shows the demographics of participants (N=9); where there exists a dash data was either unreported by the primary caregiver or not applicable.

Table 1. Vocal repertoire participant demographics.

Subject	**Age in Months**	**Gender**	**Age at ADOS Testing**	**Handedness**	**Age of First Words**	**ADOS Score**	**ADOS Speech Level**
1	98	M	98	L	-	15	SW
2	129	M	129	R	30	18	SW
3	201	F	201	R	36	22	SW
4	117	M	117	R	24	23	FNW
5	100	F	100	-	-	22	FNW
6	131	F	131	R	-	14	FNW
7	170	M	170	R	96	19	FNW
8	102	M	102	R	-	13	FNW
9	116	M	116	R	18	19	FNW

2.1.2. Annotation scheme design

An annotation scheme developed by Buder, Warlaumount and Oller (2013) characterizing the volitional and communicative vocalizations of typically developing infants was adapted for use in clinical cases as with NMVA. The original scheme reflects how infant protophones become increasingly speech-like over the first year of life before the onset of words, reaching a peak that results in articulated, syllabified and well-timed utterances that blend vocalic and closant elements. Reflexive vocalizations including burps, sneezes, coughs, hiccups and yawns were excluded from final coding as they do not provide insight into intentional communication through speechlike vocalization, but are merely uncontrolled responses to physiological functions (Kent 2022); however, they were included in the bifurcated annotation scheme in order to provide a comprehensive overview of vocalizations and guide for the annotation process. In the adapted design, certain protophones such as the "goo," "raspberry," "marginal babbling" and "canonical babbling" were renamed as quasiclosant, bilabial vibrant, marginal syllables and canonical syllables respectively, in order to reflect the biological maturity of participants past 18 months of age, the time when typically developing children have acquired word-level speech. Another protophone, "fusses," was removed entirely for the same reasoning and the grunt was moved from the reflexive to the volition category. One protophone was created for the scheme, the closant, which will be discussed in detail in the following paragraphs. Protophones were organized in descending order by post-natal range of onset, although this can vary greatly due to a myriad of factors not discussed here.

Within the principal branch of interest titled *volitional* are *phonation* and *articulation*, the latter category being further divided into *non-syllabic* and *syllabic* types. Phonated protophones are the earliest and most rudimentary vocalizations, produced by the rapid expansion and contraction of the vocal cords coupled with voicing, resulting in vocalic sound energy. While articulators modulate movement of the vocal tract and airflow unobstructed, sound can also be forced through the nasal cavity which results in nasalization, or damping.

The first vocalization within the phonation branch is the cry; it can be coupled with a sad, angry or fearful affect conveying distress (Oller et al., 2021) as with the yell, a higher pitched phonatory vocalization that appears around 4 months to 6 months of age. Quasivowels emerge shortly after a cry, around 3 months of age, and can precede or coincide with yells. This particular vocalization is produced with a neutral oral cavity whereby unimpeded airflow creates a short, quiet sound that possesses temporal aspects of adult speech (Oller 2000). Full vowels can have the same onset time as quasivowels, but are intentionally postured, sounding much like an isolated vowel, and emitted with greater force. Squeals and growls are on contrasting ends of the pitch range; squeals are twice as high as one's normal pitch register and often with excitement or other positive emotional states, whereas growls are lower than a normal pitch register with a harsh or coarse quality. The grunt, with a highly variable onset from around 3 months to 16 months old, was counted as a volition vocalization due to the communicative intent that can be conveyed, such as displeasure or a bid for attention. Bordenave and Mccune (2021) found non-verbal children with disabilities used grunts in a more communicative fashion than their typically developed peers, while children with pre-linguistic language levels (level lower than 9 months) did not produce any communicative grunts. Robb et al. (2019) also characterized an infant's exercise in laryngeal restriction, necessary to produce a grunt, as a necessary precursor to phonetic skillset. The final phonated vocalizations, the whisper and ingressive or ingressive-egresive sequence, lack vocal fold vibration and are primarily descriptive of the orientation of airflow via the lungs. Whispers are created with expanded vocal folds with agitated phonation resulting in a hushed speech sound; an ingressive sequence is gasping of air towards the lungs, as if to communicate shock or surprise; and lastly, the ingressive-egressive sequence is an in-out motion of air emanating from the lungs that can be quick or paced in timing.

Articulated vocalizations are, as the name suggests, a configuration of airflow by the articulators (lips, alveolar ridge, hard palate, soft palate, larynx, pharynx) that result in a closant sound when not syllabified, or a closant-vocalic blend when syllabified. Non-syllabic protophones include the bilabial vibrant, click, quasiclosant and closant. At around 3 months old, vocalizations such as the bilabial vibrant, formerly the raspberry, and click appear. Both protophones allow for somatosensory and proprioceptive feedback crucial for more complex speech mechanisms (Kearney and Guenther 2019). The bilabial vibrant is a trill of the lips, whereas in the click, the tongue body or tip connects with the hard palate and negative pressure is then released from the oral cavity. The quasiclosant, a goo in the original annotation scheme, and closant are parallel to the quasivowel and full vowel; the first being the more neutral and primitive consonant, while the following is stronger in phonatory energy and more clearly articulated. In the final branch, phonatory control, vocalic and closant sounds, somatosensory exploration and increasing motor skill culminate in the most advanced section of the vocal repertoire - syllabic vocalizations.

Syllabic vocalizations mark a significant milestone in an infant's vocal scheme, as there is evidence that syllabification of speech sounds not only rely on growing neural circuitry, but external learning. Caregivers may engage more with infants that produce syllabic vocalizations, as they are more speechlike in nature; this interaction signals to the infant that producing this salient vocalization type results in increased social and auditory feedback (Warlaumont and Finnegan 2016). The glottal stop sequence, which can appear at birth, is an opening and closing of the glottis that abruptly impedes phonatory airflow; this particular protophone serves as a primer for later plosive consonants. Marginal syllables arrive around 3 months to 5 months of age, and like quasivowels and quasiclosants, are an underdeveloped variant of canonical syllables that follow. Marginal syllables are composed of a CV or VC shape with a slow trackable onset and apparent transition within the syllable. The canonical syllable stage has been reached once infants produce the same shape with a quick onset and rapid transition C and V or V and C, usually around 7 to 10 months of age. Canonical syllables can be variegated, with two different syllable shapes occurring next to one another, or reduplicated, two of the same shape occurring in tandem; however, in this scheme mono and multisyllabic shapes of any form were counted as a single marginal or canonical syllable. Finally, around 12 to 18 months old children will utter their first words, or a whole phonemically understood part of speech with complete and easily understood phonetic elements in the native language. Throughout this period infants in their burgeoning vocal capacity might produce word approximations, or a combination of consonants and vowels that are apparent attempts to articulate a normal word in their native language (McCurry and Irwin 1953). The full annotation scheme used for this study can be found in Appendix A.

2.1.3. Coding criteria

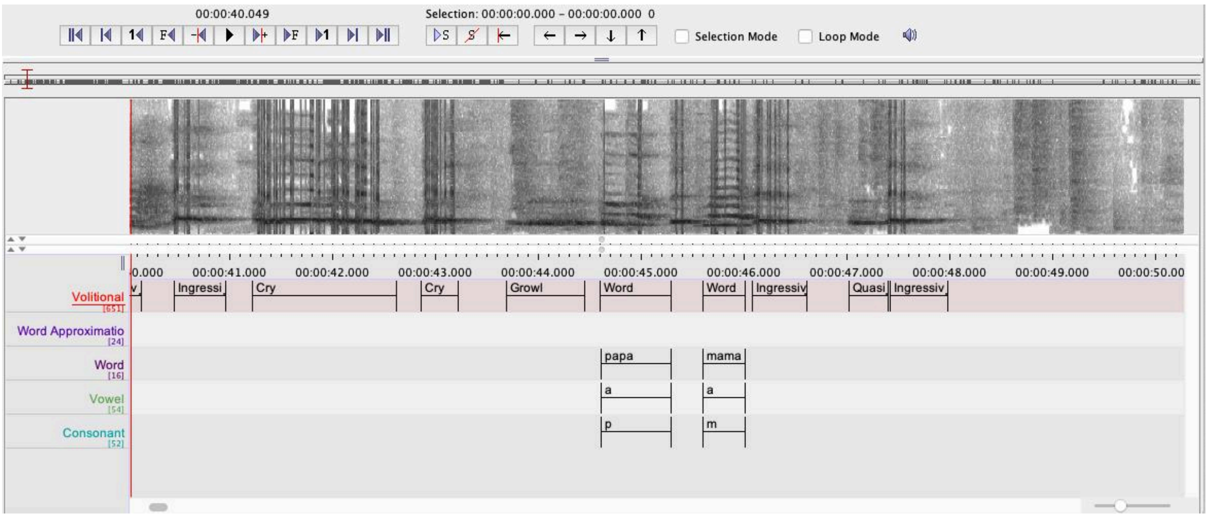
The following coding protocols were established prior to any data annotation, based on the Buder, Warlaumont and Oller's (2013) scheme. Of overarching importance was for the rater to make all coding decisions based on the participant's willingness to vocalize in a communicative manner. In some circumstances, reflexive vocalizations such as a cough, grunt or ingressive sequence can be produced with force in the absence of a biological function, which required the coder to look at communicative context to determine whether vocalizations such as these warrant coding. The rater was required to make coding judgements in line with their immediate auditory perception; that is to say if there was a conflict regarding whether a vocalization was a growl or grunt, the rater would elect to choose the vocalization type based on their first impression (Buder et al., 2013). If two protophones were uttered within the same temporal period, the more salient and complex vocalization was coded; for example, if the rater heard both a squeal and full vowel, the full vowel would be coded over the squeal as it is the more communicatively advanced vocalization. A spectrogram was relied upon for marking temporal boundaries of a vocalization, or to make a determination between two competing coding possibilities.

Other protocols were created spontaneously during the annotation process. A forced coding procedure was adopted, as many vocalizations, regardless of whether phonated or articulated, were so atypical in prosody that if auditory impression were relied on alone, the majority of subjects' vocalizations would be too difficult to classify. An unexpected vocalization that was not anticipated or included in the annotation scheme, but was present so much so that it warranted inclusion, was humming. Humming has been found to be a behavior characteristic of NMVA youths, but not those with verbal ASD (Schwartz et al., 2020) as a means of self-regulating in a noisy environment, although it can also communicate contentment and is nevertheless a vocalization. As such, it was annotated as a marginal syllable as it is syllabified and includes vocalic sounds followed by hummed nasal consonants /m/, /n/ and /ŋ/ (Gregg 2002). The singularly occurring glottal stop was also heard; however, inclusion of the glottal stop sequence required a separate coding measure for this vocalization. Due to properties such as syllabification and vocalic-closant elements, the singular glottal stop was coded as a marginal syllable as well. Finally, a note on the codification of word approximations and words. This study did not intend to analyze semantic use and comprehension, morphological competence or syntactic structure; therefore, any and all audible word approximations and words were annotated regardless of context, morphological or syntactic errors. Only clear phonological errors were annotated as word approximations, a clear example in the sample was the word "jabón" being pronounced as "jamón."

2.1.4. Materials

Nine video recordings of ADOS semi-structure diagnostic sessions used for Slušná (2021) were used for Study 1, with a mean run time of 31 minutes. Each video was saved in mp4 format and then converted into an mp3 sound only files using Audacity®: Free Audio Editor and Recorder (v3.1.3; The Audacity Team). The video and sound files were uploaded into the linguistic annotation computer software program ELAN (v6.3; Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive) for coding. A controlled vocabulary was created titled "NMVA Annotation Scheme" which included all vocalizations used in the annotation scheme (both available in the appendix). The NMVA Annotation Scheme controlled vocabulary was linked to the “Volitional” tier only, so that once a time boundary was selected the rater could double-click within the boundary and choose the vocalization from the drop-down list. Subsequent tiers included “Word Approximation,” “Word,” “Vowel” and “Consonant,” by once again marking the temporal boundaries of a new annotation, clicking and typing. Tiers and annotation examples can be seen in Figure 2.

Figure 2. Annotation screen set up in ELAN.



2.1.5. Analysis

2.1.5.1. Speechlike complexity across the group

To answer the first research question regarding the proportion of speech-like vocalizations in NMVA, normalized values were calculated by subject using the total number of each category of interest, resulting in a positive or negative percentage out of 100. For example, to determine the percentage of more speechlike syllabic vocalizations in comparison with less speechlike non-syllabic vocalizations, the following formula was used:

$$100 \times \left(\frac{\text{Total syllabic vocalizations (x)} - \text{Total nonsyllabic vocalizations(y)}}{\text{Total nonsyllabic vocalizations (y)} + \text{Total syllabic vocalizations(x)}}\right)$$

This formula was used to obtain the production proportions for the following less speechlike to more speechlike vocalizations: phonation versus articulation, non-syllabic versus syllabic, marginal syllables versus canonical syllables and word approximations versus words. Normalized values were also checked for normality visually and by performing a Shapiro-Wilk test (phonation/articulation = $p > 0.71$; syllabic/non-syllabic = $p < .001$; marginal/canonical = $p < .04$; word approximation/word = $p < .03$). In order to see if a significant difference existed from the hypothesized median of 0, the one sample Wilcoxon Signed-Rank test was performed on non-parametric data. One-sample t-tests were performed on parametric data.

Next, to address the second questions of whether quasivowels and canonical syllables independently correlated with the total words and word approximations a subject produced, the total count of words and word approximations was input for each subject. Normality of total quasivowels and total canonical syllables were assessed visually and via a Shapiro-Wilk test (quasivowels = $p < .03$; canonical syllables = $p < .004$). A Pearson's correlation was run for age and total vocalizations to observe whether age might be an influence on the total number of vocalizations produced, thereby confounding further analysis; the correlation was not significant ($p > .64$). Since both data points were not normally distributed and a scatterplot of the data showed a monotonic relationship, a Spearman's correlation was selected.

2.1.5.2. Vocal repertoire by subject

A vocal repertoire including types, tokens, vowels and consonants, was created for each subject. Types and tokens were calculated manually within Microsoft Excel. Vowels and consonants were also compiled into an inventory manually based on the types and tokens uttered by each participant. A ratio calculation was done on each vowel and consonant to see the percentage each was produced out of all 9 subjects; vowels and consonants were first listed individually and then by height (for vowels) and manner of articulation (for consonants). Importantly, this study did not seek to NMVA youths' phonetic repertoire, which would require a vowel or consonant to be present in the same position (initial, middle or final) in at least two different words (Stoel-Gammon 1985). Nor could it be reliably assessed whether NMVA individuals produce stable phonemes, as then the issue of semantic competence would arise which was not within the scope of this study. Therefore, any type or token and the vowels and/or consonants present in them were added to the inventory.

Statistical analyses were run in IBM SPSS Statistics Version 28.0, and all Shapiro-Wilk tests and violin plots (ggplot 2; Wickham 2016) were run in RStudio (RStudio Team 2022).

2.2. Results

2.2.1. Speechlike vocalizations

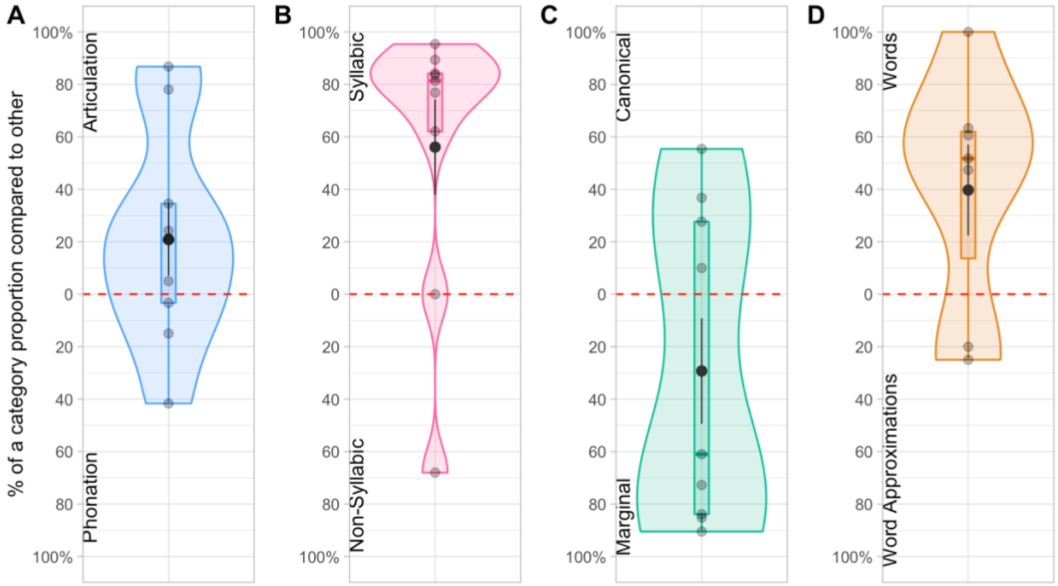
Table 2 summarize the results of all statistical tests that were run in order to answer the primary research question of proportions of speechlike complexity in NMVA. It was observed that phonation did not differ significantly from articulation, t(8) = 1.51, *p* = .169. In terms of non-syllabic and syllabic production, the group made significantly more syllabic than non-syllabic vocalizations, *z* = 2.24, *p* = .025; there was no statistically significant difference between marginal and canonical syllable production, *z* = -1.48, *p* = .14. The comparison of the proportions of word approximation and words, only reached marginal significance, t(8) = 2.13, *p* = .06. Figure 3 shows plots A-D which represent the percentage of one category produced in relation to the other. For the correlation analysis, results revealed that quasivowels, a foundational pre-speech vocalization, did not correlate significantly with total word approximations and words produced, $r_s(7) = -1.26$, *p* = .74; however, canonical syllables, a speech-like vocalization that occurs later in the first year of life, were significantly correlated with total word approximations plus words produced, $r_s(7) = .69$, *p* = .037.

Table 2. Results of the proportion and correlation analyses by test type and test item.

Proportion Analysis		**One Sample T-Test**			**CI of the Difference**		**Effect Sizes**		
Mean ± Standard Deviation		t	Two-sided p	Mean Difference	95% CI Lower	95% CI Upper	Cohen's d	95% CI Lower	95% CI Upper
Phonation Articulation	20.98 ± 41.60	1.51	0.16	20.98	-10.99	52.96	0.504	-0.206	1.18
Word Approximations Words	30.88 ± 43.35	2.13	0.06	30.88	-2.43	64.21	0.71	-0.043	1.43
One Sample Wilcoxon Signed-Rank Test									
		Standard Error			Sig.	z	Median		
Non-syllabic Syllabic		7.14			.025*	2.24	81.36		
Marginal Syllables Canonical Syllables		8.44			0.14	-1.48	60.98		

Note. * indicates a significant result at p < .05. Spearman's correlation is omitted from the table as all values are reported in the text.

Figure 3. Distribution of proportions produced of one vocalization type in relation to another. Mean + SE are shown.



Note. The y-axis is measured in positive or negative values out of 100 with 0 meaning equal proportions of two variables. Plot A. displays Phonation and Articulation; Plot B. displays Non-Syllabic and Syllabic; Plot C. displays Marginal Syllables and Canonical Syllables; Plot D. displays Word Approximations and Words.

2.2.2. Vocal repertoire

Individual and group results of the vocal repertoire revealed that out of all the volitional vocalizations in the group (*N* = 2,585), only 8.47% were words and 3.4% were word approximations. Words also made up a majority of total tokens of word approximations and words produced for all subjects (*N* = 308*)* at 71.10%. Only Subject 1 and Subject 8 produced more word approximations than words; Subjects 5 and 6 produced no word approximations or words at all. Looking at the variation in production of word approximations and words across the group, 70.78% of word approximations tokens were of various types and word types made up 54.79% of all word tokens.

All 5 vowels and 20 out of 21 consonants were represented within the entire group. The high vowel /i/ and mid vowel /e/ had the lowest occurrence, yet still present in over 50% of the entire group. The consonant /x/ was absent, and 9 consonants did not reach the 50% threshold: /c/, /d/, /f/, k/, /q/, /r/, /v/, /w/ and /z/. Sorted by manner of articulation, this amounts to 4 plosives, 3 fricatives, 1 liquid and 1 glide. The lowest produced consonants, only reaching 22% of the group, were the plosive /k/, fricative /f/ and fricative /v/; at 33% were plosive /d/, fricative /z/ and glide /w/ and finally at 44% were plosive /c/, plosive /q/ and liquid /r/. Table 3 shows the vocal repertoire for each subject (*N* = 9), including total word approximation tokens and types, total word tokens and types as well as vowels and consonants.

Table 3. Vocal repertoire by subject.

Subject	**Word Approximation Tokens**	**Word Approximation Types**	**Word Tokens**	**Word Types**	**Vowels**	**Consonants**
1	24	9	16	7	a, o, u	b, g, h, j, k, m, n, p, s, t, w, y
2	12	10	53	34	a, e , i, o, u	b, c, d, f, g, j, l, m, n, p, q, r, s, t v, y, z
3	19	10	66	20	a, e , i, o, u	b, c, d, g ,h, j, l, m, n, p, q, r, s, t, y, z
4	0	0	1	1	i, o	l, s , t
5	0	0	0	0	-	-
6	0	0	0	0	-	-
7	24	24	66	45	a, e, i, o, u	b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, y, z
8	5	5	3	3	a,e ,o, u	b, h, j, m, n, w, y
9	5	5	14	10	a, e, i, o, u	b, c, g, h, l, m, n, p, q, r, s, t, y

3. Study 2: Relation of vocal repertoires to white matter integrity of the AF

3.1. Methods

3.1.1. Participants

Participants were eight (*N* = 8) of the nine youths who participated in the first above study on vocal repertoires. Subject 5 did not return for MRI imagine and was therefore excluded from this portion; Subject 6 was diagnosed with co-morbid 22q11 deletion syndrome. Table 4 shows demographics of each participant.

Table 4. Participant demographics for MRI imaging.

Subject	**Age in Months**	**Gender**	**Handedness**
1	99	M	L
2	133	M	R
3	212	F	R
4	113	M	R
-	-	-	-
6	147	F	R
7	184	M	R
8	114	M	R
9	133	M	R

3.1.2. Diffusion MRI: data acquisition, preprocessing and estimation

Diffusion MRI data were acquired for each participant using a 3T Ingenia CX scanner (Philips Medical Systems) located at the Hospital Sant Juan de Déu (Barcelona, Spain) with a standard 32-channel head coil and the following sequence parameters: Field-of-view = 230x230 mm; voxel-size = 2.05x2.05mm2; repetition time (TR) = 10.1s; echo time (TE) = 102ms; flip angle = 90º; number-of-slices = 64; slice-thickness = 2.1mm; number of averages = 1; acceleration factor = 2; number of shells = 2; b-values = 625, and 1250 s/mm^2; number of diffusion gradient directions = 36; number of b0 (i.e., b-value=0) images = 3, plus 1 b0 with reverse phase to correct for spatial distortions. For each subject, the diffusion MRI data were corrected to remove susceptibility, eddy-current, and head motion distortions using the topup (Andersson et al., 2003) and eddy (Andersson & Sotiropoulos, 2016) toolboxes in FSL (https://fsl.fmrib.ox.ac.uk/fsl/) (Smith et al., 2004). The FA metric was obtained from the computed diffusion tensors, which were estimated by using the *dtifit* program included in FSL.

3.1.3. Multi-Echo T2 Relaxometry: data acquisition, preprocessing and estimation

Multi-echo T2 (MET2) data were acquired for all participants during the same scanning session described in the previous section by using a gradient and spin-echo sequence (Prasloski et al., 2012b) with the following parameters: Field-of-view = 240 × 230 mm; acquisition voxel-size = 3.5 × 3.5 mm^2 , reconstructed voxel-size = 1.6 × 1.6 mm ; number of echoes = 32; echo-time spacing (ΔTE) = 8.38 ms; repetition time (TR) = 8620 ms; excitation pulse = 90^0; refocusing pulses = 1800; number-of-slices = 40; slice-thickness = 3.5 mm; number of averages = 1; acceleration factor (SENSE) = 2; acquisition time = 8:37 min.

The intra-voxel T2 distribution of relaxation times was calculated by using regularized non-negative least squares (Laule et al., 2007; Mackay & Laule, 2012) with a regularization term to promote smooth solutions that better represent the distribution expected from tissue microstructure (Mackay et al., 1994; Whittall et al., 1997) as described in (Guo et al., 2013). The estimation was carried out using the open-source multi-component T2 reconstruction toolbox (Canales-Rodríguez et al., 2021a; Canales-Rodríguez, Alonso-Lana, et al., 2021b; Canales-Rodríguez, Pizzolato, et al., 2021c) available at https://github.com/ejcanalesr/multicomponent-T2-toolbox. The implementation is based on the extended phase graph (EPG) model (Prasloski, Mädler, et al., 2012) and uses a T2 discrete grid from 10-2000ms (Prasloski, Rauscher, et al., 2012) with p=60 T2 logarithmically spaced points.

From the estimated T2 distributions, the MWF was calculated as the area under the curve for T2 times smaller than the myelin water cutoff T2=40ms, normalized by the total area under the curve of the whole T2 distribution (Meyers et al., 2017).

3.1.4. Tract segmentation

Bundle-specific tractograms were generated using TractSeg (Wasserthal et al., 2018), an automatic machine learning algorithm that segments the WM into 72 bundles (https://github.com/MIC-DKFZ/TractSeg). It generates tract masks as well as tract orientation maps which enable the creation of accurate bundle-specific tractograms. For each subject, the mean FA and MWF values over the tract masks were estimated for the following language-related tracts: the arcuate fasciculus (AF).

3.1.5. Analysis of FA and MWF in the AF

In order to answer the question of how vocal measures from the NMVA repertoire map onto the microstructure of the left and right AF, the analysis included two different metrics of white matter integrity, namely FA and MWF. Total articulation and total phonation were selected as the vocal measures of interest to map onto the AF metrics, which also related to the overarching question of speechlike complexity in NMVA. Additionally, utilizing these two main tiers from the coding scheme removed any confound in statistical testing, as other nested vocalization types may actually correlate or be predictive of one another.

Fractional anisotropy (FA) is a Diffusion Tensor Imaging (DTI) metric based on the axonal diffusion of water molecules which serves as an informative marker of fiber organization and proxy of myelination (Johansen-Berg and Behrens 2013). FA is measured from 0 to 1; lower values signify highly isotropic diffusion and therefore lowered white matter integrity, while higher values indicate anisotropic diffusion corresponding with greater white matter integrity (DiProspero et al., 2022). A multiple regression model was fit to predict whether total phonation and total articulation predicted FA values of the left AF, and then a second multiple regression was run with the same regressors, with the outcome variable changed to FA of the right AF. Taking into consideration Slušná (2021; see also Kochunov et al., 2012) multi-tract discovery of FA values increasing with age in NMVA, including in the left and right AF, age was added as an additional regressor in the models.

Myelin Water Fraction (MWF) is a measure of myelin water to total water; the former being defined as water between myelin bilayers (Laule et al., 2007). Correlation analyses were performed with total phonation and total articulation with MWF values of the left and right AF respectively. WMF values did not show any correlation with age in NMVA in the same sample (Slušná 2021), and thus were not included in any correlation analysis. A Spearman's correlation was selected as linearity and normality could not be visually assessed with confidence and total articulation was not normally distributed; therefore, the assumptions needed to fit a regression model were not met.

The Shapiro-Wilk test was run in to assess the normality of total phonation ($p = .05$) and total articulation ($p = .001$). A Spearman's correlation for non-normally distributed data was performed on total phonation and total articulation to assess whether there would exist a confound in conducting multiple regressions on the FA values and vocal repertoire measures. The correlation was not statistically significant, $r_s(7) = .310$, $p = .417$, hence no confound would limit running the multiple regression models for FA.

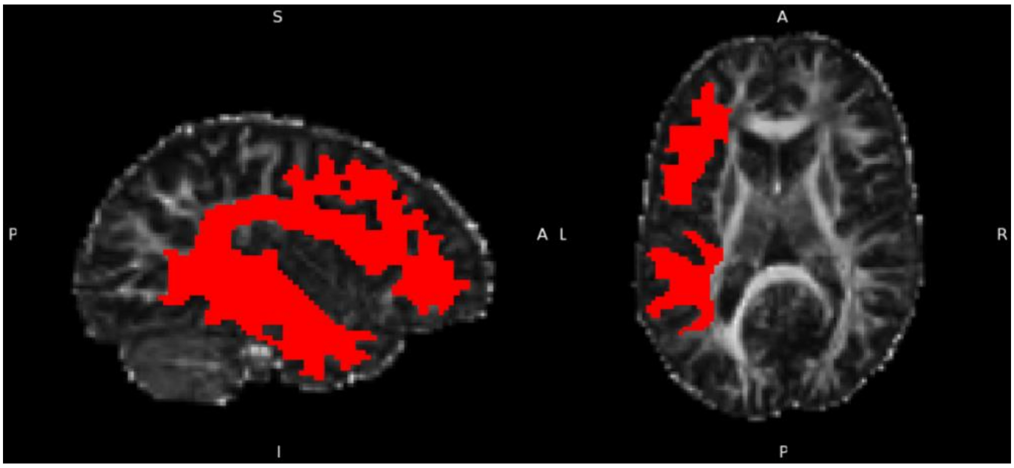
The Spearman's correlation and multiple linear regression models were run in IBM SPSS Statistics Version 28.0 and Shapiro-Wilk tests in RStudio (RStudio Team 2022).

3.2. Results of FA and MWF in the AF

For the left AF, the regressors were not significant predictors of FA values, *F*(3, 4) = 3.069, *p* = .153, adjusted R^2 = .470. The model was significant for age only in the right AF, *F*(3, 4) = 10.639, *p* = .022, adjusted R^2 = .805.

The Spearman's correlation for the MWF values did not reveal significant correlations between total articulation and the MWF of the left AF, $r_s(7) = -.023$, p = .978, or the MWF of the right AF, $r_s(7) = -.395$, p = .333. Nor did total phonation significant correlate with the MWF of the left AF, $r_s(7) = .467$, p = .132, or the MWF of the right AF, $r_s(7) = .243$, p = .756.

Figure 4. FA masks over a T_2 image of the brain showing a sagittal view of the left AF (left) and axial view of the bilateral AF (right).



4. Discussion

The first portion of this thesis set out to characterize the vocal repertoire of individuals with NMVA, guided by an annotation scheme originally developed for neurotypical infants. In summary, results revealed that subjects produced more speechlike than non-speechlike vocalizations. Thus, when looking at pairs of distinctive non-speechlike and speechlike variables (phonation vs. articulation, non-syllabic vs. syllabic, marginal vs. canonical and word approximations vs. words), the non-speechlike vocalizations did not surpass speechlike vocalizations in *any* comparison. In fact, two speechlike vocalizations outweighed their non-speechlike counterparts with full statistical significance, as with non-syllabic vs. syllabic vocalizations, or marginal statistical significance, as with word approximations vs. words. Words were verbalized more frequently at over 70% of the summation of word approximations and words produced across the entire sample.

This preliminary finding has foundational implications for how NMVA is conceptualized. If the NMVA speech trajectory was fundamentally thwarted by dysarthria or CAS, it might be expected that these individuals vocalize largely in phonated protophones like quasivowels, and that these might correlate with word approximations and words, but that was not seen here. Additionally, if dysarthria or CAS was an explanatory factor in NMVA, individuals might even produce more articulated non-syllabic vocalizations, which was also not substantiated in this study. While phonated vocalizations were still present in the repertoire at a later age than expected, they were secondary in comparison with more speech-like vocalizations including marginal and canonical syllables. Proportions of marginal and canonical syllables did not reach significance, and word approximations vs. words only reached marginal significance, revealing that across the group aspects of neurotypical and atypical development continue to comingle alongside one another. In neurotypical development these vocalizations become more scaffolded and as a more complex vocalization type appears, the other transitions out of the repertoire. In particular, marginal syllables appear and subtlety decline upon the appearance of canonical syllables, which then begin to phase out as word approximations and words come into the linguistic picture.

That canonical syllables correlated with word approximations and words is an aspect shared with neurotypical vocal development, whose onset and ratio of canonical babbling has been found to be a predictor of lexical development and size in neurotypical infants (Cychosz et al., 2021). It must still be contextualized that out of the total vocalizations produced (N = 2,585) in the entire sample, only 8.47% were words and 3.4% were word approximations.

The individual vocal repertoires provide further evidence to this emerging depiction of NMVA. A relative stability of vowels was seen across the group, with only the high vowel /i/ and the mid vowel /e/ failing to reach 50% of production at group level. This aligns in part with neurotypical development, whereby low unrounded vowels are produced more frequently in the first year of life for neurotypical infants (McLeod and Bleile 2003). As with vowels, hallmarks of neurotypical consonant acquisition were observed as well; fricatives and liquids were only produced in 22% of the group here and have been found to be later acquired in neurotypical children's repertoire: liquids at ~5 years 11 months and fricatives at ~6 years 11 months old (McLeod and Crowe 2018). Plosives were unusual here, as they were also produced across 22% of the group, but are typically acquired early in neurotypical development, ~3 years 11 months (McLeod and Crowe 2018).

We were unable to find compelling evidence of a relationship between vocal variation in NMVA and a speech-related tract, the AF. Vocalization variables including total phonation and total articulation did not predict or correlate with two white matter microstructural metrics in either hemisphere of the AF. It is worth noting previous research using FA found atypical lateralization in the AF within NMVA youths of the same age (Slušná 2021), and as such we expected that total articulation would be a predictor of FA in the right AF; however, based on the present findings, we can conclude that although the AF displays anomalies in NMVA, the variation in vocalization repertoires does not relate to FA in the AF. Results for vocal measures were not significantly correlate with MWF of the left or right AF; however, there was moderate indirect correlation observed between total articulation and MWF of the right AF, and a moderate direct correlation between total phonation and MWF of the left AF. Interestingly, the same multi-tract study by Slušná (2021) found that MWF of the bilateral AF did not increase with age in NMVA youths, although the opposite has been documented in neurotypical infants, children and adults (Deoni et al., 2012; Dvorak et al., 2021). Warlaumont and Finnegan (2016) illustrated canonical syllables, a syllabic articulation, are not present at birth and rely not only on biological and neurological maturation, but for neural training to occur via the infant's modulation of caregiver talk into structured syllables that include a vowel and consonant. This would offer specific support for a prediction of significance in the left AF's MWF, insofar as myelination is supported throughout the lifespan by the continuous use and adaptation of cognitive functions such as speech; yet it was not found here.

The totality of our results raises further questions as to variation in the vocalization profile of NMVA and how such variation was not found to be related to a key speech-related tract (Broce et al., 2015). We surmise that previously purported predictors of expressive language in NMVA, such as non-verbal IQ (Saul and Norbury 2020) and non-verbal cognition (Yoder et al., 2015), were perhaps responsible for participants' increasing or decreasing speechlike vocalizations. If this is case, it could stand to reason subjects with higher non-verbal IQ or non-verbal cognition produced more speechlike vocalizations. Moreover, these speechlike vocalizations may not be linguistic in nature, in the sense that the type of vocalizations we assessed, articulation and phonation, do not carry coherent semantic meaning as with neurotypical vocal production. This could also be an explanation of why these vocal measures did not result in significant findings when analyzed with two microstructural metrics of the AF, a crucial part of the neural speech network (Catani et al., 2005). Nonetheless, it cannot be settled that all vocalizations in NMVA are independent of the AF, but this body of work points to growing evidence of abnormalities in of the language connectome within NMVA (Olivé et al., 2022; Slušná 2021).

Our results were obtained in a sample spanning a wider age range of individuals with NMVA, including school-aged children and teens (Saul et al., 2020; Chenausky et al., 2017; Chenausky et al., 2018). Previous studies of NMVA have utilized other vocal measures such as imitated speech, or relied solely on clinical assessment, with subjects whom had undergone some form of speech therapy, in order to investigate white matter integrity in the AF (Chenausky et al., 2017; Wan et al., 2012). The use of non-naturalistic and restrictive speech sampling may not be fully representative of the biology of the AF and how it may or may not be altered in NMVA. By contrast to such sampling, the second above study analyzed subject's spontaneous speech in a flexible, play-based setting, which can provide important complementary information.

There is also much more to be done in the way of vowel and consonant inventories in NMVA, which form the basis of semantic content and have been found to be predictive of neurotypical children's later vocabulary size. Extensive speech sampling is dependent on this work and was outside the possible timeframe of this study. Further speech sampling and a larger sample size would also provide the statistical power needed to confirm whether the moderate correlations observed between articulation and phonation, would then reach significance with MWF of the right AF and left AF respectively. Furthermore, vocal repertoire measures need to be mapped onto other language-related tracts that have yet to be explored in detail within NMVA, such as the IFOF and UF, which are ventral stream tracts highlighted in Olivé et al., (2022).

4.1. Limitations

This study relied on a small sample size (*N* = 9 for vocal repertoires; *N* = 8 for Diffusion MRI) and therefore lacked statistical power. Recruitment in special populations, as is the case with NMVA, is difficult and therefore much of the previous research that has inspired and been cited in this study has had to rely upon small sample sizes in order to begin exploring this disorder. Moreover, the investigation into the vocal repertoire of NMVA was annotated by a single rater, ergo no inter-rater reliability was conducted. Any results obtained here can only serve as an exploratory foundation on which future work can be directed, and excitingly so.

5. Conclusion

This study proved to be the first in-depth analysis of vocalizations in NMVA, revealing the existence of speech-like complexity even in a non- or minimally verbal population. While other explanatory factors such as dysarthria or CAS cannot be eliminated as contributing factors to the behavioral phenotype of this disorder, it is remarkable that speechlike vocalizations are not universally beyond this population's capacity. Future work needs to replicate the vocal repertoire study in a larger sample in order to corroborate this finding, while subsequently mapping variation in the vocal repertoires to the entire structural language connectome using a multi-metric microstructural approach, as was done here. The last word on myelin-related measures in NMVA is thus not yet said; and it is important to understand which variation in the linguistic phenotype involved, if not in vocalization, variation in myelination relates to.

6. References

Almairac, F., Herbet, G., Moritz-Gasser, S., de Champfleur, N. M., & Duffau, H. (2015). The left inferior fronto-occipital fasciculus subserves language semantics: A multilevel lesion study. *Brain Structure and Function*, *220*(4), 1983–1995.

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019

Audacity Team. (n.d.). *Audacity(R): Free Audio Editor and Recorder* (3.1.3) [Computer software]. https://audacityteam.org/

Bachorowshi, J.-A., & Owren, M. J. (2004). Psychological Science Agenda| September 2004. *Psychological Science*.

Bal, V. H., Maye, M., Salzman, E., Huerta, M., Pepa, L., Risi, S., & Lord, C. (2020). The adapted ADOS: A new module set for the assessment of minimally verbal adolescents and adults. *Journal of Autism and Developmental Disorders*, *50*(3), 719–729.

Bedoya, S., Katz, N. B., Brian, J., O'Shaughnessy, D., & Falk, T. H. (2020). 3. Acoustic and prosodic analysis of vocalizations of 18-month-old toddlers with autism spectrum disorder. In A. Neustein & H. A. Patil (Eds.), *Acoustic Analysis of Pathologies: From Infancy to Young Adulthood* (pp. 93–126). De Gruyter. https://doi.org/doi:10.1515/9781501513138-003

Bordenave, D., & Mccune, L. (2021). Grunt Vocalizations in Children With Disabilities: Relationships With Assessed Cognition and Language. *Journal of Speech, Language, and Hearing Research*, *64*, 1–11. https://doi.org/10.1044/2021_JSLHR-21-00202

Broce, I., Bernal, B., Altman, N., Tremblay, P., & Dick, A. S. (2015). Fiber tracking of the frontal aslant tract and subcomponents of the arcuate fasciculus in 5–8-year-olds: Relation to speech and language function. *Brain and Language*, *149*, 66–76. https://doi.org/10.1016/j.bandl.2015.06.006

Canales-Rodríguez, E. J., Alonso-Lana, S., Verdolini, N., Sarró, S., Feria, I., Montoro, I., Garcia-Ruiz, B., Jimenez, E., Varo, C., & Albacete, A. (2021). Age-and gender-related differences in brain tissue microstructure revealed by multi-component T2 relaxometry. *Neurobiology of Aging*, *106*, 68–79.

Canales-Rodríguez, E. J., Pizzolato, M., Piredda, G. F., Hilbert, T., Kunz, N., Pot, C., Yu, T., Salvador, R., Pomarol-Clotet, E., & Kober, T. (2021). Comparison of non-parametric T2 relaxometry methods for myelin water quantification. *Medical Image Analysis*, *69*, 101959.

Canales-Rodríguez, E. J., Pizzolato, M., Yu, T., Piredda, G. F., Hilbert, T., Radua, J., Kober, T., & Thiran, J.-P. (2021). Revisiting the T2 spectrum imaging inverse problem: Bayesian regularized non-negative least squares. *NeuroImage*, *244*, 118582.

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*.

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C. A., Thompson, C. K., Schotten, M. T. de, Dell'Acqua, F., Weintraub, S., & Rogalski, E. J. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain : A Journal of Neurology*, *136 Pt 8*, 2619–2628.

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Special Issue on "Brain Hodology - Revisiting Disconnection Approaches to Disorders of Cognitive Function," 44*(8), 1105–1132. https://doi.org/10.1016/j.cortex.2008.05.004

Chenausky, K., Brignell, A., Morgan, A., & Tager-Flusberg, H. (2019). Motor speech impairment predicts expressive language in minimally verbal, but not low verbal, individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *4*, 2396941519856333. https://doi.org/10.1177/2396941519856333

Chenausky, K., Kernbach, J., Norton, A., & Schlaug, G. (2017). White Matter Integrity and Treatment-Based Change in Speech Performance in Minimally Verbal Children with Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00175

Chenausky, K. V., Norton, A. C., Tager-Flusberg, H., & Schlaug, G. (2018). Behavioral predictors of improved speech output in minimally verbal children with autism. *Autism Research*, *11*.

Chenausky Karen V., Brignell Amanda, Morgan Angela T., Norton Andrea C., Tager-Flusberg Helen B., & Schlaug Gottfried. (2021). A Modeling-Guided Case Study of Disordered Speech in Minimally Verbal Children With Autism Spectrum Disorder. *American Journal of Speech-Language Pathology*, *30*(3S), 1542–1557. https://doi.org/10.1044/2021_AJSLP-20-00121

Childhood Apraxia of Speech [Technical Report]. (2007). American Speech-Language-Hearing Association. https://www.asha.org/policy/TR2007-00278/#r49

Choi, B., Leech, K. A., Tager-Flusberg, H., & Nelson, C. A. (2018). Development of fine motor skills is associated with expressive language outcomes in infants at high and low risk for autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, *10*.

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, *24*. https://doi.org/10.1111/desc.13090

Davis, B. L., & Velleman, S. L. (2000). Differential diagnosis and treatment of developmental apraxia of speech in infants and toddlers. *Infant-Toddler Intervention: The Transdisciplinary Journal*, *10*, 177–192.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, *298*, 2013–2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., & Dubois, J. (2006). Nature and nurture in language acquisition: Anatomical and functional brain-imaging studies in infants. *Trends in Neurosciences*, *29*(7), 367–373. https://doi.org/10.1016/j.tins.2006.05.011

Deoni, S. C. L., Dean, D. C., O'Muircheartaigh, J., Dirks, H., & Jerskey, B. A. (2012). Investigating white matter development in infancy and early childhood using myelin water faction and relaxation time mapping. *NeuroImage*, *63*(3), 1038–1053. https://doi.org/10.1016/j.neuroimage.2012.07.037

Deoni, S. C. L., Dean, D. C., Remer, J., Dirks, H., & O'Muircheartaigh, J. (2015). Cortical maturation and myelination in healthy toddlers and young children. *NeuroImage*, *115*, 147–161. https://doi.org/10.1016/j.neuroimage.2015.04.058

DeVeney, S. L., Kyvelidou, A., & Mather, P. (2021). A home-based longitudinal study of vocalization behaviors across infants at low and elevated risk of autism. *Autism & Developmental Language Impairments*, *6*, 23969415211057656. https://doi.org/10.1177/23969415211057658

Dick, A., Bernal, B., & Tremblay, P. (2013). The Language Connectome: New Pathways, New Concepts. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *20*. https://doi.org/10.1177/1073858413513502

DiProspero, N. D., Kim, S., & Yassa, M. A. (2022). Chapter 8—Magnetic resonance imaging biomarkers for cognitive decline in Down syndrome. In E. Head & I. Lott (Eds.), *The*

Neurobiology of Aging and Alzheimer Disease in Down Syndrome (pp. 149–172). Academic Press. https://doi.org/10.1016/B978-0-12-818845-3.00014-1

Draganova, R., Eswaran, H., Murphy, P., Lowery, C., & Preissl, H. (2007a). Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Human Development*, *83*(3), 199–207. https://doi.org/10.1016/j.earlhumdev.2006.05.018

Draganova, R., Eswaran, H., Murphy, P., Lowery, C., & Preissl, H. (2007b). Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Human Development*, *83*(3), 199–207. https://doi.org/10.1016/j.earlhumdev.2006.05.018

Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Hüppi, P. S., & Hertz-Pannier, L. (2014). The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. *Neuroscience*, *276*, 48–71. https://doi.org/10.1016/j.neuroscience.2013.12.044

Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J. F., Bihan, D. L., & Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensori-motor networks. *Cerebral Cortex*, *19 2*, 414–423.

Dvorak, A. V., Swift-LaPointe, T., Vavasour, I. M., Lee, L. E., Abel, S., Russell-Schulz, B., Graf, C., Wurl, A., Liu, H., Laule, C., Li, D. K. B., Traboulsee, A., Tam, R., Boyd, L. A., MacKay, A. L., & Kolind, S. H. (2021). An atlas for human brain myelin content throughout the adult life span. *Scientific Reports*, *11*(1), 269. https://doi.org/10.1038/s41598-020-79540-3

Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *The Journal of Pediatrics*, *124*(2), 199–203. https://doi.org/10.1016/S0022-3476(94)70303-5

Friederici, A. D. (2009). Pathways to language: Fiber tracts in the human brain. *Trends in Cognitive Sciences*, *13*(4), 175–181. https://doi.org/10.1016/j.tics.2009.01.001

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal Exposures to Sound and Vibroacoustic Stimulation. *Journal of Perinatology*, *20*(1), S21–S30. https://doi.org/10.1038/sj.jp.7200446

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, *19*, 123–137.

Gregg, J. W. (2002). Vocal development and articulation in speech and song. *Journal of Singing*, *58*(5), 431–438.

Guo, J., Ji, Q., & Reddick, W. E. (2013). Multi-slice myelin water imaging for practical clinical applications at 3.0 T. *Magnetic Resonance in Medicine*, *70*(3), 813–822.

Hepper, P. G., & Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *71*(2), F81–F87. https://doi.org/10.1136/fn.71.2.F81

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.

Hickok, G., & Poeppel, D. (2007). The Cortical Organization of Speech Processing. *Nature Reviews. Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113

Hinzen, W., Slusna, D., Schroeder, K., Sevilla, G., & Borrellas, E. (2019). Mind–Language =? The significance of non-verbal autism. *Mind & Language*, *35*. https://doi.org/10.1111/mila.12257

Hus, V., Pickles, A., Cook, E. H., Risi, S., & Lord, C. (2007). Using the Autism Diagnostic Interview—Revised to Increase Phenotypic Homogeneity in Genetic Studies of Autism. *Biological Psychiatry*, *61*(4), 438–448. https://doi.org/10.1016/j.biopsych.2006.08.044

Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders*, *54*, 32–42. https://doi.org/10.1016/j.jcomdis.2015.01.002

Johansen-Berg, H., & Behrens, T. E. (2013). *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press.

Kearney, E., & Guenther, F. (2019). Articulating: The Neural Mechanisms of Speech Production. *Language, Cognition and Neuroscience*, *34*, 1–16. https://doi.org/10.1080/23273798.2019.1589541

Kent, R. (2022). The maturational gradient of infant vocalizations: Developmental stages and functional modules. *Infant Behavior and Development*, *66*, 101682. https://doi.org/10.1016/j.infbeh.2021.101682

Kochunov, P., Williamson, D. E., Lancaster, J., Fox, P., Cornell, J., Blangero, J., & Glahn, D. C. (2012). Fractional anisotropy of water diffusion in cerebral white matter across the lifespan. *Neurobiology of Aging*, *33*(1), 9–20. https://doi.org/10.1016/j.neurobiolaging.2010.01.014

Koegel, L. K., Bryan, K. M., Su, P. L., Vaidya, M., & Camarata, S. (2020). Definitions of Nonverbal and Minimally Verbal in Research for Autism: A Systematic Review of the Literature. *Journal of Autism and Developmental Disorders*, *50*(8), 2957–2972. https://doi.org/10.1007/s10803-020-04402-w

La Valle, C., Chenausky, K., & Tager-Flusberg, H. (2021). How do minimally verbal children and adolescents with autism spectrum disorder use communicative gestures to complement their spoken language abilities? *Autism & Developmental Language Impairments*, *6*, 239694152110350. https://doi.org/10.1177/23969415211035065

Laake, L. M., & Bridgett, D. J. (2014). Happy babies, chatty toddlers: Infant positive affect facilitates early expressive, but not receptive language. *Infant Behavior and Development*, *37*(1), 29–32. https://doi.org/10.1016/j.infbeh.2013.12.006

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical Babbling: A Marker for Earlier Identification of Late Detected Developmental Disorders? *Current Developmental Disorders Reports*, *6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

Laule, C., Vavasour, I. M., Kolind, S. H., Li, D. K., Traboulsee, T. L., Moore, G. W., & MacKay, A. L. (2007). Magnetic resonance imaging of myelin. *Neurotherapeutics*, *4*(3), 460–484.

Lebersfeld, J., Swanson, M., Clesi, C., & O'Kelley, S. (2021a). Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children. *Journal of Autism and Developmental Disorders*, *51*. https://doi.org/10.1007/s10803-020-04839-z

Lebersfeld, J., Swanson, M., Clesi, C., & O'Kelley, S. (2021b). Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children. *Journal of Autism and Developmental Disorders*, *51*. https://doi.org/10.1007/s10803-020-04839-z

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. E. (2010). Autism Spectrum Disorder and Co-occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States. *Journal of Developmental & Behavioral Pediatrics*, *31*(4). https://journals.lww.com/jrnldbp/Fulltext/2010/05000/Autism_Spectrum_Disorder_and_Co_occurring.1.aspx

Linke, A. C., Slušná, D., Kohli, J. S., Álvarez-Linera Prado, J., Müller, R.-A., & Hinzen, W. (2021). Morphometry and functional connectivity of auditory cortex in school-age children with profound language disabilities: Five comparative case studies. *Brain and Cognition*, *155*, 105822. https://doi.org/10.1016/j.bandc.2021.105822

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*, *284*.

Lynch, M. P., Oller, D. K., Steffens, M. L., & Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental Psychobiology*, *28 1*, 3–25.

MacKay, A. L., & Laule, C. (2007). Myelin water imaging. *EMagRes*.

Mackay, A., Whittall, K., Adler, J., Li, D., Paty, D., & Graeb, D. (1994). In vivo visualization of myelin water in brain by magnetic resonance. *Magnetic Resonance in Medicine*, *31*(6), 673–677.

Maenner, M., Shaw, K., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D., Wiggins, L., Pettygrove, S., Andrews, J., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C., Lee, L.-C., Harrington, R., Huston, M., & Dietz, P. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR. Surveillance Summaries*, *69*, 1–12. https://doi.org/10.15585/mmwr.ss6904a1

Mahmoudzadeh Mahdi, Dehaene-Lambertz Ghislaine, Fournier Marc, Kongolo Guy, Goudjil Sabrina, Dubois Jessica, Grebe Reinhard, & Wallois Fabrice. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences*, *110*(12), 4846–4851. https://doi.org/10.1073/pnas.1212220110

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, *19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Marschik, P., Pokorny, F., Peharz, R., Zhang, D., O'Muircheartaigh, J., Roeyers, H., Bölte, S., Spittle, A., Urlesberger, B., Schuller, B., Poustka, L., Ozonoff, S., Pernkopf, F., Pock, T., Tammimies, K., Enzinger, C., Krieber, M., Tomantschger, I., Bartl-Pokorny, K., & Group, T. (2017). A Novel Way to Measure and Predict Development: A Heuristic Approach to Facilitate the Early Detection of Neurodevelopmental Disorders. *Current Neurology and Neuroscience Reports*, *17*, 43. https://doi.org/10.1007/s11910-017-0748-8

Mccune, L. (2021). Laryngeal vocalizations: Grunts as a gateway to language? *Lingua*, *264*, 103161. https://doi.org/10.1016/j.lingua.2021.103161

McCurry William H. & Irwin Orvis C. (1953). A Study Of Word Approximations In The Spontaneous Speech Of Infants. *Journal of Speech and Hearing Disorders*, *18*(2), 133–139. https://doi.org/10.1044/jshd.1802.133

McDaniel, J., Slaboch, K. D., & Yoder, P. (2018). A meta-analysis of the association between vocalizations and expressive language in children with autism spectrum disorder. *Research in Developmental Disabilities*, *72*, 202–213. https://doi.org/10.1016/j.ridd.2017.11.010

McLeod, S., & Bleile, K. (2003). *Neurological and developmental foundations of speech acquisition*. 1–12.

McLeod Sharynne & Crowe Kathryn. (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

Meyers, S. M., Kolind, S. H., & MacKay, A. L. (2017). Simultaneous measurement of total water content and myelin water fraction in brain at 3 T using a T2 relaxation based method. *Magnetic Resonance Imaging*, *37*, 187–194.

Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., & Bryson, S. (2006). Early Language and Communication Development of Infants Later Diagnosed with Autism Spectrum Disorder. *Journal of Developmental & Behavioral*

Pediatrics, *27*(2). https://journals.lww.com/jrnldbp/Fulltext/2006/04002/Early_Language_and_Communication_Development_of.4.aspx

Mody, M., Shui, A. M., Nowinski, L. A., Golas, S. B., Ferrone, C., O'Rourke, J. A., & McDougle, C. J. (2017). Communication Deficits and the Motor System: Exploring Patterns of Associations in Autism Spectrum Disorder (ASD). *Journal of Autism and Developmental Disorders*, *47*, 155–162.

Nathani, S., & Oller, D. K. (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 321–330. https://doi.org/10.3758/BF03195385

Nijmegen. (n.d.). *ELAN (Version 6.3) [Computer software]*. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., Gillberg, C., & Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, *19*(8), 934–943. https://doi.org/10.1177/1362361314556782

Olivé, G., Slušná, D., Vaquero, L., Muchart-López, J., Rodríguez-Fornells, A., & Hinzen, W. (2022). Structural connectivity in ventral language pathways characterizes non-verbal autism. *Brain Structure and Function*, *227*(5), 1817–1829. https://doi.org/10.1007/s00429-022-02474-1

Oller, D. K. (2000). *The emergence of the speech capacity*.

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C.-C., Bowman, D. D., Long, H. L., Buder, E. H., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports*, *9*(1), 14734. https://doi.org/10.1038/s41598-019-51352-0

Oller, D. K., Ramsay, G., Bene, E., Long, H. L., & Griebel, U. (2021). *Protophones, the precursors to speech, dominate the human infant vocal landscape*.

O'Muircheartaigh, J., Dean, D. C., 3rd, Ginestet, C. E., Walker, L., Waskiewicz, N., Lehman, K., Dirks, H., Piryatinsky, I., & Deoni, S. C. L. (2014). White matter development and early cognition in babies and toddlers. *Human Brain Mapping*, *35*(9), 4475–4487. PubMed. https://doi.org/10.1002/hbm.22488

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal Patterns in Infants with Autism Spectrum Disorder: Canonical Babbling Status and Vocalization Frequency. *Journal of Autism and Developmental Disorders*, *44*(10), 2413–2428. https://doi.org/10.1007/s10803-014-2047-4

Peter, B. (2013). Biological substrates of speech development: A brief synopsis of the developing neuromuscular system. In *Comprehensive Perspectives on Speech Sound Development and Disorders: Pathways from Linguistic Theory to Clinical Practice* (pp. 79–101). Nova Science Publishers, Inc.

Peter, B., Macleod, A., Buder, E., Warlaumont, A., & Oller, D. K. (2013). *In: Comprehensive Perspectives … Editors: AN ACOUSTIC PHONETIC CATALOG OF PRESPEECH VOCALIZATIONS FROM A DEVELOPMENTAL PERSPECTIVE*.

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech Acquisition in Older Nonverbal Individuals With Autism: A Review of Features, Methods, and Prognosis. *Cognitive and Behavioral Neurology*, *22*(1). https://journals.lww.com/cogbehavneurol/Fulltext/2009/03000/Speech_Acquisition_in_Older_Nonverbal_Individuals.1.aspx

Plate, S., Yankowitz, L., Resorla, L., Swanson, M., Meera, S., Estes, A. M., Marrus, N., Cola, M., Petrulla, V., Faggen, A., Pandey, J., Paterson, S., Pruett, J., Cody, H., Dager, S., John, T., Botteron, K., Zwaigenbaum, L., Piven, J., & Parish-Morris, J. (2021). Infant vocalizing and phenotypic outcomes in autism: Evidence from the first 2 years. *Child Development*, *93*. https://doi.org/10.1111/cdev.13697

Prasloski, T., Mädler, B., Xiang, Q., MacKay, A., & Jones, C. (2012). Applications of stimulated echo correction to multicomponent T2 analysis. *Magnetic Resonance in Medicine*, *67*(6), 1803–1814.

Prasloski, T., Rauscher, A., MacKay, A. L., Hodgson, M., Vavasour, I. M., Laule, C., & Mädler, B. (2012). Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence. *Neuroimage*, *63*(1), 533–539.

Pujol, J., Soriano-Mas, C., Ortiz, H., Sebastián-Gallés, N., Losilla, J. M., & Deus, J. (2006). Myelination of language-related areas in the developing brain. *Neurology*, *66*(3), 339. https://doi.org/10.1212/01.wnl.0000201049.66073.8d

Rapin, I., Dunn, M., Allen, D. A., Stevens, M., & Fein, D. (2009). *Subtypes of language disorders in school-age children with autism (Developmental Neuropsychology (34, 1, (66-84)). 34*. https://doi.org/10.1080/87565640902769295

Robb Michael P., Yavarzadeh Faraz, Schluter Philip J., Voit Verena, Shehata-Dieler Wafaa, & Wermke Kathleen. (2020). Laryngeal Constriction Phenomena in Infant Vocalizations. *Journal of Speech, Language, and Hearing Research*, *63*(1), 49–58. https://doi.org/10.1044/2019_JSLHR-S-19-0205

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Yendiki, A., Rowe, M. L., & Gabrieli, J. D. E. (2018). Language Exposure Relates to Structural Neural Connectivity in Childhood. *The Journal of Neuroscience*, *38*, 7870–7877.

RStudio Team. (n.d.). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, *29*(2003), 30.

Samson, A. C., Dougherty, R. F., Lee, I. A., Phillips, J. M., Gross, J. J., & Hardan, A. Y. (2016). White matter structure in the uncinate fasciculus: Implications for socio-affective deficits in Autism Spectrum Disorder. *Psychiatry Research: Neuroimaging*, *255*, 66–74.

Saul, J., & Norbury, C. (2020). Does phonetic repertoire in minimally verbal autistic preschoolers predict the severity of later expressive language impairment? *Autism*, *24*(5), 1217–1231. https://doi.org/10.1177/1362361319898560

Saur Dorothee, Kreher Björn W., Schnell Susanne, Kümmerer Dorothee, Kellmeyer Philipp, Vry Magnus-Sebastian, Umarova Roza, Musso Mariacristina, Glauche Volkmar, Abel Stefanie, Huber Walter, Rijntjes Michel, Hennig Jürgen, & Weiller Cornelius. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*(46), 18035–18040. https://doi.org/10.1073/pnas.0805234105

Saygin, Z. M., Norton, E. S., Osher, D. E., Beach, S. D., Cyr, A. B., Ozernov-Palchik, O., Yendiki, A., Fischl, B. R., Gaab, N., & Gabrieli, J. D. E. (2013). Tracking the Roots of Reading Ability: White Matter Volume and Integrity Correlate with Phonological Awareness in Prereading and Early-Reading Kindergarten Children. *The Journal of Neuroscience*, *33*, 13251–13258.

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research*, *4*.

Schwartz, S., Shinn-Cunningham, B., & Tager-Flusberg, H. (2018). Meta-analysis and systematic review of the literature characterizing auditory mismatch negativity in individuals with autism. *Neuroscience & Biobehavioral Reviews*, *87*, 106–117. https://doi.org/10.1016/j.neubiorev.2018.01.008

Schwartz, S., Wang, L., Shinn-Cunningham, B., & Tager-Flusberg, H. (2020). Atypical Perception of Sounds in Minimally and Low Verbal Children and Adolescents With Autism as Revealed by Behavioral and Neural Measures. *Autism Research*, *13*. https://doi.org/10.1002/aur.2363

Sheinkopf, S. J., Mundy, P. C., Oller, D. K., & Steffens, M. L. (2000). Vocal Atypicalities of Preverbal Autistic Children. *Journal of Autism and Developmental Disorders*, *30*, 345–354.

Slušná, D. (2021). *Lower-level auditory processing of speech in nvASD*. Pompeu Fabra University.

Stoel-Gammon Carol. (1985). Phonetic Inventories, 15–24 Months. *Journal of Speech, Language, and Hearing Research*, *28*(4), 505–512. https://doi.org/10.1044/jshr.2804.505

Su, P., Kuan, C.-C., Kaga, K., Sano, M., & Mima, K. (2008). Myelination progression in language-correlated regions in brain of normal children determined by quantitative MRI assessment. *International Journal of Pediatric Otorhinolaryngology*, *72*(12), 1751–1763. https://doi.org/10.1016/j.ijporl.2008.05.017

Tager-Flusberg, H., Rogers, S. J., Cooper, J., Landa, R., Lord, C., Paul, R., Rice, M. L., Stoel-Gammon, C., Wetherby, A. M., & Yoder, P. J. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research : JSLHR*, *52 3*, 643–652.

Tierney, C., Mayes, S., Lohs, S. R., Black, A., Gisin, E., & Veglia, M. (2015). How Valid Is the Checklist for Autism Spectrum Disorder When a Child Has Apraxia of Speech? *Journal of Developmental and Behavioral Pediatrics : JDBP*, *36*(8), 569—574. https://doi.org/10.1097/dbp.0000000000000189

Wan, C. Y., Marchina, S., Norton, A., & Schlaug, G. (2012). Atypical hemispheric asymmetry in the arcuate fasciculus of completely nonverbal children with autism. *Annals of the New York Academy of Sciences*, *1252*(1), 332–337.

Warlaumont, A. S., & Finnegan, M. K. (2016). Learning to produce syllabic speech sounds via reward-modulated neural plasticity. *PloS One*, *11*(1), e0145096.

Wasserthal, J., Neher, P., & Maier-Hein, K. H. (2018). TractSeg—Fast and accurate white matter tract segmentation. *NeuroImage*, *183*, 239–253. https://doi.org/10.1016/j.neuroimage.2018.07.070

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werwach, A., Mürbe, D., Schaadt, G., & Männel, C. (2021). Infants' vocalizations at 6 months predict their productive vocabulary at one year. *Infant Behavior and Development*, *64*, 101588. https://doi.org/10.1016/j.infbeh.2021.101588

Whittall, K. P., Mackay, A. L., Graeb, D. A., Nugent, R. A., Li, D. K., & Paty, D. W. (1997). In vivo measurement of T2 distributions and water contents in normal human brain. *Magnetic Resonance in Medicine*, *37*(1), 34–43.

Wickham, Hadley. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Willadsen, E., Persson, C., Patrick, K., Lohmander, A., & Oller, D. K. (2020). Assessment of prelinguistic vocalizations in real time: A comparison with phonetic transcription and assessment of inter-coder-reliability. *Clinical Linguistics & Phonetics*, *34*(7), 593–616. https://doi.org/10.1080/02699206.2019.1681516

Williamson, J. M., & Lyons, D. A. (2018). Myelin Dynamics Throughout Life: An Ever-Changing Landscape? *Frontiers in Cellular Neuroscience*, *12*. https://www.frontiersin.org/article/10.3389/fncel.2018.00424

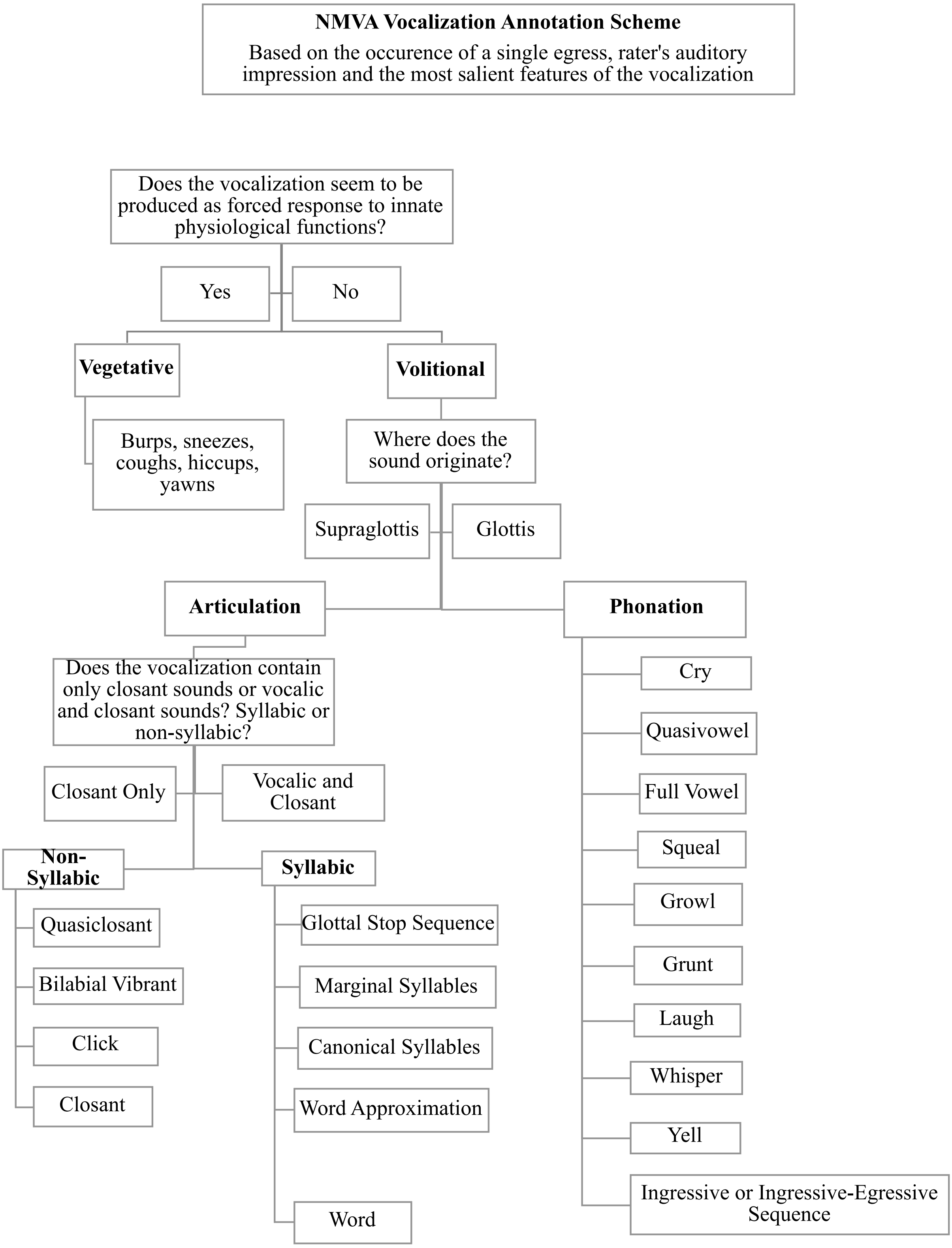
Yankowitz, L., Schultz, R., & Parish-Morris, J. (2019). Pre- and Paralinguistic Vocal Production in ASD: Birth Through School Age. *Current Psychiatry Reports*, *21*. https://doi.org/10.1007/s11920-019-1113-1

Yoder, P., Stone, W. L., Walden, T., & Malesa, E. (2009). Predicting Social Impairment and ASD Diagnosis in Younger Siblings of Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *39*(10), 1381–1391. https://doi.org/10.1007/s10803-009-0753-0

Yoder, P., Watson, L. R., & Lambert, W. (2015). Value-Added Predictors of Expressive and Receptive Language Growth in Initially Nonverbal Preschoolers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(5), 1254–1270. https://doi.org/10.1007/s10803-014-2286-4

Zuk, J., Yu, X., Sanfilippo, J., Figuccio, M. J., Dunstan, J., Carruthers, C., Sideridis, G. D., Gagoski, B. A., Grant, P. E., & Gaab, N. (2019). White matter in infancy is prospectively associated with language outcome in kindergarten. *BioRxiv*.

7. Appendix A


NMVA Vocalization Annotation Scheme
Based on the occurence of a single egress, rater's auditory impression and the most salient features of the vocalization
Does the vocalization seem to be produced as forced response to innate physiological functions?
Yes
No
Vegetative
Volitional
Burps, sneezes, coughs, hiccups, yawns
Where does the sound originate?
Supraglottis
Glottis
Articulation
Phonation
Does the vocalization contain only closant sounds or vocalic and closant sounds? Syllabic or non-syllabic?
Closant Only
Vocalic and Closant
Non-Syllabic
Syllabic
Quasiclosant
Bilabial Vibrant
Click
Closant
Glottal Stop Sequence
Marginal Syllables
Canonical Syllables
Word Approximation
Word
Cry
Quasivowel
Full Vowel
Squeal
Growl
Grunt
Laugh
Whisper
Yell
Ingressive or Ingressive-Egressive Sequence

Annotation Manual for Vocalizations

Adapted from Buder, Warlaumont and Oller (2013)

Phonation

Phonation based vocalizations are voiced and considered vocalic (vowel) in nature. Vocal cords are set into vibration primarily by the glottis and differ from syllabification/articulation related vocalizations as they are not syllabified by any articulatory mechanisms. The articulators serve to mold the oral and pharyngeal cavities into the require position for vocalic sounds and also push airflow through the nasal cavity (Peter 2013), which can then lead to a nasal quality in the vocalization.

Cry (from birth): A cry is counted as a single occurrence in this scheme, even if syllabified. Cries are the first vibration of the vocal cords in healthy newborns, after the lungs take their first inhale and exhale of air. In linguistic development, cries reflect the prosodic contours of the native/ambient language that are transmitted prenatally and acquired by the fetus (Wermke et al., 2009). A vocalic sound, crying also communicates affective states to caretakers including sadness, frustration and anger.

Quasivowel (~3 months): The quasivowel is a foundational vocalic sound that is made as the result of egressive airflow producing unforced phonation with the vocal tract in a semi-relaxed state. This vocalization can be defined by its short duration, low sound energy and less mature/developed production.

Full Vowel (~3 months): A full vowel is well-defined, with higher sound energy and longer in duration. This vocalization is auditorily more mature with a vocal tract shaped in a way that produces a more developed and clear utterance.

Squeal (~3 months): Squeals are high in pitch and considered a functionally flexible vocalization (Oller et al. 2019), often associated with excitement. They differ from a yell in that they are not stressed or forced, but rather with a vocal tract that maintains a steadiness at twice the level of the individuals normal pitch range.

Growl (~3 months): The growl consists of a hoarse, abrasive sound quality that goes below the individual's normal pitch range.

Grunt (~3-16 months): Grunts are vocalic vocalizations that are made due to physical exertion, focused attention or with communicative intent (McCune 2021). Grunts originating due to physical effort are annotated here, as they are a sign of respiratory and laryngeal fitness – while in some cases also cuing others for assistance. The grunt sound is made with an open vocal tract and restriction of the laryngeal articulator, which in healthy infants is the first articulatory apparatus to be controlled prior to developing their phonetic skills (Robb et al., 2019).

Laugh (~4 months): Laughter is an unarticulated vocalic vocalization most often made in a positive emotional state in social contexts. Whether a laugh is produced in one egress or multi-syllabic, the instance is annotated as a single occurrence. Laughter can be voiced, whereby the vocal cords vibrate in a melodic fashion, or unvoiced, which can be perceived as a grunt especially in males (Bachorowshi & Owen 2004).

Whisper (~4-6 months): A whisper is an egressive vocalization whereby the vocal folds of the larynx are abducted so that they cannot vibrate.

Yell (~4-6 months): A yell is a stressed or forced sound that is high above the individuals' normal pitch range with a functional flexibility that can express fear, excitement, anger, etc.

Ingressive or Ingressive-Egressive Sequence (~3-8 months): Ingressive sounds are made with an inhalation of air towards the lungs – as opposed to egressive sounds. Ingressive sounds can be made in sequence and become syllabified, but are marked as a single occurrence. These sounds can be made to emote fear, surprise or shock. Ingressive-eggressive sequences are an inhalation and exhalation of air from the lungs that can also be made sequentially in rapid succession.

Syllabification/Articulation

Closant (consonant) sounds that are produced when the articulators constrict airflow within the pharynx and subsequently within the vocal tract via planned and coordinated motor movements. These vocalizations can be syllabified by supraglottal phonation, but also by the glottis. The glottal stop sequence is included here, in the *syllabification/articulation* type of vocalization, as the end result is a *syllabified* vocalization, unlike phonated vocalizations.

Bilabial Vibrant (~3 months): The lips come together and air is expelled rapidly cause the lips to vibrate in rapid succession. This type of vocalization is highly sensory and signals the individual is exploring their vocal capacity.

Click (~3-8 months): The tongue body makes contact with the hard (and occasionally soft) palate in a negatively pressurized environment within the oral cavity. This sound emitted is primarily articulated and lacks voicing (phonation).

Quasiclosant (~1-4 months): (Replacing "goo") The quasi-closant replaces the "goo," a defining vocalization of infants and their developing ability for speech. The articulation occurs primarily in the posterior portion of the oral cavity, but lacks the well-formedness and organized structure of a fully developed consonant sound. A quasiclosant instead contains more variable features, forming what could be considered a pre-consonant.

Closant (~5+ months): A full, well-formed consonant occurring *without* an accompanying vowel.

Glottal Stop Sequence (from birth): Voicing produced within the larynx whereby the phonatory pattern of airflow oscillates from interrupted to uninterrupted. This articulatory related vocalization is a potential precursor to plosives as control is needed over egressive airflow to create this manner of consonant. The vocalization is categorized under *syllabification* as the sequential egressive airflow creates syllable-like timing that is quick with identifiable boundaries.

Marginal Syllables (~3-5 months): Marginal syllables are supraglottal articulations with a slow onset time and a CV or VC shape. Dependent on the syllable shape, this particular syllable type can lack a clear nucleus. A marginal syllable can be distinguished by the transcriber as being a more advanced vocalization than ones previously mentioned, yet falls short of the rapid onset, articulatory complexity and mature timing in the transition from C to V or V to C.

Canonical Syllables (~7-10 months): The canonical syllable can be easily identified by the transcriber as having a smooth, rapid transition from C to V or V to C and a quicker onset. The canonical syllable can be clearly perceived by its more mature posture in the vocal tract and with a distinguishable nucleus. This syllable type can either be reduplicated (such as "baba") or variegated (such as "mani").

Word Approximation: A word approximation is an attempt to produce a complete part of speech, but is thwarted by an inserted, omitted or mispronounced phonetic element.

Word (~12-18 months): A complete part of speech and existing phonetic elements in an individual's native language. A word must be easily identified and annotated with near perfect agreement amongst transcribers. Each word is sub-annotated into its vowels and consonants.