

Assessment of Childhood Apraxia of Speech: A Review/Tutorial of Objective Measurement Techniques

Hayo Terband,^a Aravind Namasivayam,^b Edwin Maas,^c Frits van Brenk,^d Marja-Liisa Mailend,^e Sanne Diepeveen,^{f,g} Pascal van Lieshout,^b and Ben Maassen^h

Background: With respect to the clinical criteria for diagnosing *childhood apraxia of speech* (commonly defined as a disorder of speech motor planning and/or programming), research has made important progress in recent years. Three segmental and suprasegmental speech characteristics—error inconsistency, lengthened and disrupted coarticulation, and inappropriate prosody—have gained wide acceptance in the literature for purposes of participant selection. However, little research has sought to empirically test the diagnostic validity of these features. One major obstacle to such empirical study is the fact that none of these features is stated in operationalized terms.

Purpose: This tutorial provides a structured overview of perceptual, acoustic, and articulatory measurement procedures that have been used or could be used to

operationalize and assess these 3 core characteristics. Methodological details are reviewed for each procedure, along with a short overview of research results reported in the literature.

Conclusion: The 3 types of measurement procedures should be seen as complementary. Some characteristics are better suited to be described at the perceptual level (especially phonemic errors and prosody), others at the acoustic level (especially phonetic distortions, coarticulation, and prosody), and still others at the kinematic level (especially coarticulation, stability, and gestural coordination). The type of data collected determines, to a large extent, the interpretation that can be given regarding the underlying deficit. Comprehensive studies are needed that include more than 1 diagnostic feature and more than 1 type of measurement procedure.

^aUtrecht Institute of Linguistics-OTS, Utrecht University, the Netherlands

^bOral Dynamics Laboratory, Department of Speech-Language Pathology, University of Toronto, Ontario, Canada

^cDepartment of Communication Sciences and Disorders, Temple University, Philadelphia, PA

^dDepartment of Communicative Disorders and Sciences, University at Buffalo, NY

^eMoss Rehabilitation Research Institute, Moss Rehabilitation Hospital, Elkins Park, PA

^fHAN University of Applied Sciences, Nijmegen, the Netherlands

^gDepartment of Rehabilitation, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands

^hCenter for Language and Cognition, Research School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands

Correspondence to Hayo Terband: h.r.terband@uu.nl

Editor-in-Chief: Julie Liss

Received May 11, 2019

Accepted May 18, 2019

https://doi.org/10.1044/2019_JSLHR-S-CSMC7-19-0214

Publisher Note: This article is part of the Special Issue: Select Papers From the 7th International Conference on Speech Motor Control.

From a historical perspective, childhood apraxia of speech (CAS) is a controversial clinical entity, with respect to both clinical signs and underlying deficit. In 1981, Guyette and Diedrich had concluded that "...No pathognomonic symptoms or necessary and sufficient conditions were found for the diagnosis..." (p. 44) and critically termed CAS as "a label in search of a population" (p. 39). Despite clinical studies to further characterize CAS (e.g., Aram & Horwitz, 1983; Ekelman & Aram, 1984; Marion, Sussman, & Marquardt, 1993; Pollock & Hall, 1991; B. Smith, Marquardt, Cannito, & Davis, 1994; Walton & Pollock, 1993), this situation had not changed much by the time of 1994, when Shriberg (1994) concluded that development in this field was moving endlessly sideways.

Since then, a large body of research has been dedicated to characterize the speech impairment and underlying functional and neuromotor deficit of CAS, and this endeavor has been successful in some respects. There is an agreement that, from a functional point of view, CAS is a disorder of motor planning and/or motor programming (American Speech-Language-Hearing Association [ASHA], 2007) or,

Disclosure: The authors have declared that no competing interests existed at the time of publication.

in other words, an inability to transform an abstract phonological code into motor speech commands (cf. Maassen, Nijland, & Terband, 2010). More specifically, ASHA defined CAS as “a neurological childhood (pediatric) speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits (e.g., abnormal reflexes, abnormal tone)... The core impairment in planning and/or programming spatiotemporal parameters of movement sequences results in errors in speech sound production and prosody.” (ASHA, 2007, pp. 3–4). Since then, this definition has been adopted widely in the CAS research literature (e.g., Grigos & Kolenda, 2010; Iuzzini-Seigel, Hogan, Guarino, & Green, 2015; Maas & Farinella, 2012; Murray, McCabe, Heard, & Ballard, 2015; Namasivayam et al., 2015; Preston et al., 2014; Terband, Maassen, Guenther, & Brumberg, 2009, 2014).

With respect to the clinical criteria for diagnosing CAS, research has also made important progress in recent years. Although ASHA (2007, p. 4) noted that “there is no validated list of diagnostic features of CAS that differentiates this symptom complex from other types of childhood speech sound disorders,” the CAS Technical Report proposed three segmental and suprasegmental speech characteristics that were considered to be consistent with a deficit in speech motor planning and programming and thus as being specific to CAS:

1. inconsistent errors on consonants and vowels in repeated productions of syllables or words;
2. lengthened and disrupted coarticulatory transitions between sounds and syllables; and
3. inappropriate prosody, especially in the realization of lexical or phrasal stress.

These features have gained wide acceptance in the subsequent literature for purposes of participant selection, but little research has sought to empirically test the diagnostic validity of these features. One major obstacle to such empirical study is the fact that none of these proposed features was stated in operationalized terms. This lack of operationalization also hinders comparability of participants across studies, because often researchers either do not provide operationalized criteria for the CAS diagnoses of their participants or researchers use different criteria. The purpose of this tutorial is to provide a structured overview of measurement procedures that have been used or could be used to operationalize and assess these three core characteristic. The hope is that this will facilitate a more replicable evidence base and, eventually, a consensus on how best to capture these features for future research and clinical application.

To be clear, we do not address whether a “feature checklist” is ultimately the optimal approach to diagnosis (e.g., see Shriberg et al., 2017, for a discussion of problems with this approach), nor do we suggest that these specific features are the most important or discriminative ones (see Murray, Iuzzini-Seigel, Maas, Terband, & Ballard,

2018, for a systematic review of the differential diagnostic value of these features). Alternative approaches such as developing psycholinguistic profiles derived from process-oriented diagnostics have been proposed elsewhere (e.g., Terband, Maassen, & Maas, 2016, 2019). The goal of the current article is to provide a structured overview of measurement procedures that have been used or may be used to assess the three core characteristics of CAS as formulated in the ASHA Technical Report (ASHA, 2007), without going into the issue of differential diagnosis itself.

This review is organized by each feature characterizing CAS and within each feature by level of analysis (perceptual/transcription, acoustic, articulatory analysis). We review methodological details for each procedure and provide a short overview of research results that have been reported in the literature. In terms of methodological details, for each approach, we identify four critical parameters that must be specified for operationalization and determining cutoff scores for diagnosis: (a) the response target to be produced by the child (sounds, words, nonwords, etc.), (b) the task used to elicit these responses (e.g., imitation, picture naming), (c) the conditions under which the responses are elicited (e.g., quiet, with time pressure), and (d) the measures obtained from these responses (e.g., error consistency scores, formant ratios). For each method, we further summarize the scientific basis, specifically, (e) whether administration is standardized, (f) whether validity and reliability data are available, and (g) whether norm or reference data for children are available (we make a distinction between *norm* data, i.e., norm-referenced cutoff scores, and *reference* data, i.e., numbers reported by other studies that may serve as reference values). Finally, we discuss issues that need to be taken into consideration when choosing a suitable technique and identify research needs in terms of the development of (more objective) measures as well as their validation and standardization.

Inconsistent Errors on Consonants and Vowels in Repeated Productions of Syllables or Words

Background

Inconsistency of Speech

Disordered or atypical “inconsistency” is variability in speech production in the absence of contextual variations (e.g., phonetic context, pragmatic influences, maturation or cognitive–linguistic influences), such as during repeated productions of the same exemplar across multiple trials (Dodd, Hua, Crosbie, Holm, & Ozanne, 2009; Marquardt, Jacks, & Davis, 2004). The measurement of inconsistent speech production includes not just quantity of different productions and control of context but also the quality of those alterations. Qualitative differences, such as the number and type of (multiple) substitutes for phonemes within and across all positions, assist in the differentiation of atypical/disordered “inconsistency” from “normal” variability as found in typically developing (TD) children (Iuzzini-Seigel, 2012; Iuzzini-Seigel & Forrest, 2010). In the

next sections, we will discuss measures that allow us to distinguish variability that is a part of normal learning and development from atypical inconsistency seen in children with speech disorder (e.g., CAS).

Speech Variability During Typical Development

In TD children, some degree of variability in word production is expected, but highly inconsistent speech production is considered a sign of pathology or disorder (Holm, Crosbie, & Dodd, 2007). In repeated productions of the same word in a picture-naming task (with 25 items), Holm et al. (2007) found approximately 10%–13% variability at the whole-word level in TD children ages 3;0–6;11 (years;months). Studies of typical speech development have documented decreasing variability during the repeated productions of words or speech sounds with increasing age (Iuzzini-Seigel, 2012; Preston & Koenig, 2011). For example, Burt, Holm, and Dodd (1999) and Holm et al. (2007) found a negative correlation between age and word variability in children with typical speech development between 3;10 and 4;10 and between 3;0 and 6;11, respectively. Within this general trend of decreasing word variability in TD children, variability peaks have been observed during certain phases, such as during language and vocabulary expansion (Iuzzini-Seigel, Hogan, Rong, & Green, 2015; Sosa & Stoel-Gammon, 2006). Specifically, Sosa and Stoel-Gammon (2006) observed an increase in whole-word variability in children between 1 and 2 years of age when two-word combinations were emerging and when vocabulary size was approximately 150–200 words. Vocabulary expansion between 15 and 21 months has also been associated with a temporary regression in speech motor performance (Iuzzini-Seigel, Hogan, Rong, et al., 2015). These nonmonotonic changes in error variability during typical development have been attributed to resource allocation issues and dynamic interactions between language and speech systems (Green, Nip, & Maassen, 2010; Iuzzini-Seigel, Hogan, Rong, et al., 2015; Macrae, Tyler, & Lewis, 2014). Overall, children's speech production is more variable, less flexible, and less accurate than adult speech until the early teens (A. Smith & Zelaznik, 2004).

Error Inconsistency in CAS

In general, studies provide evidence for increased variability in speech production of children with CAS relative to TD children or those with other speech impairments (e.g., Dodd, Hua, Crosbie, Holm, & Ozanne, 2002; Iuzzini-Seigel, Hogan, & Green, 2017; Schumacher, McNeil, Vetter, & Yoder, 1986). For example, Schumacher et al. (1986) found that whole-word phonetic variability elicited from repetitions of words distinguished children (5–9 years of age) with CAS from TD children or those with functional articulation disorders. However, results from word-level inconsistency measures (e.g., Token-to-Token Inconsistency; Dodd et al., 2002) should be interpreted cautiously. Children with inconsistent phonological disorder and children with severe speech sound disorder (SSD), in general, may demonstrate high scores on word-level inconsistency assessments,

possibly implying that word-level inconsistency may relate to the severity of the problem and not just disorder classification (Bradford & Dodd, 1996; Iuzzini-Seigel, 2012; Tyler, Williams, & Lewis, 2006). In fact, a recent study demonstrated that inconsistency scores alone (from the Diagnostic Evaluation of Articulation and Phonology [DEAP] Inconsistency subtest; Dodd et al., 2002) were only able to discriminate CAS from other SSDs with a modest accuracy of 30% (Murray et al., 2015) and thus may not be sufficient for differential diagnosis (Bradford & Dodd, 1996).

Segmental-level inconsistency measures (e.g., type-token ratio [TTR]; Forrest & Iuzzini-Seigel, 2008; Iuzzini-Seigel & Forrest, 2010) have proven to be more sensitive than word-level procedures for differential diagnosis of CAS from other SSD populations. In particular, segmental-level TTR measures, the consonant substitute inconsistency percentage (CSIP; Forrest & Iuzzini-Seigel, 2008; Iuzzini-Seigel & Forrest, 2010) and its variant, the inconsistency severity percentage (ISP; Iuzzini-Seigel & Forrest, 2010), demonstrate high scores for children with CAS but not TD children or children with articulation or phonological delays (Forrest & Iuzzini-Seigel, 2008; Iuzzini-Seigel, 2012; Yao-Tresguerres, Iuzzini-Seigel, & Forrest, 2009). For example, CSIP scores below 21% were found for children with phonological or articulatory disorders, while children with CAS had CSIP scores of greater than 24% (Forrest & Iuzzini-Seigel, 2008). Similarly, ISP scores differentiated TD children from speakers with speech disorder, with > 18% ISP scores indicating possible CAS diagnosis (TD group had ISP scores of < 7.5%). Overall, Iuzzini-Seigel (2012) suggests that between segmental (e.g., ISP) and lexical (Word Inconsistency Measure; DEAP subtest) inconsistency measures, the segmental-level analysis may be relatively more sensitive for differential diagnosis between TD, phonological disorder (PD), and CAS and to track intervention-related changes over time.

At the level of acoustic inconsistency, measures such as the acoustic spatiotemporal variability indices (e.g., envelope-based spatiotemporal index [E-STI]; Howell, Anderson, Bartrip, & Bailey, 2009) or voice onset time (VOT) variability (Iuzzini-Seigel, 2012) have clinical potential for differential diagnosis and treatment progress monitoring in CAS, but they have rarely been applied in this population. Generally, children's VOTs are more variable than adults' VOTs, and variability decreases with age and stabilizes around the age of 11 years (Auzou et al., 2000; Whiteside, Dobbin, & Henry, 2003). Iuzzini-Seigel (2012) investigated inconsistency of speech in 3- to 5-year-old children with CAS, PD, and TD using acoustic (VOT variability), segmental, and lexical measures. Children with CAS evidenced less stability at both the acoustic level (significantly higher coefficients of variation [COVs] of VOTs for bilabial voiceless stops) and at the segmental and lexical levels relative to speakers with PD and TD speakers. Furthermore, Iuzzini-Seigel also analyzed VOT measures (e.g., COV and skewness) as a function of group, differentiated by segmental (e.g., CSIP, ISP) or lexical inconsistency (e.g., Word Inconsistency Assessment; Dodd et al., 2009) measures. Only in

groups classified by the segmental-level inconsistency measures (and not groups differentiated by lexical-level inconsistency measures) did speakers with CAS demonstrate a more positive skewness, that is, a higher COV for VOTs relative to speakers with PD. In a more recent study, Iuzzini-Seigel, Hogan, Guarino, et al. (2015) demonstrated that, under conditions of attenuated auditory feedback (auditory masking), children with CAS produced a lower percentage of optimal exemplars of voiceless bilabial stops and reduced vowel space area relative to TD children or children with speech delays. They interpreted these findings as indicative of poor feedforward motor programs and compensatory reliance on auditory feedback in CAS (Terband & Maassen, 2010).

At the level of kinematic inconsistency (e.g., kinematic STI; Kleinow & Smith, 2000), studies have indicated that speech articulation is more variable in preschool- and school-age children with CAS, relative to children with other SSDs or TD peers (Grigos, Moss, & Lu, 2015; Moss & Grigos, 2012; Terband, Maassen, van Lieshout, & Nijland, 2011). For example, Grigos et al. (2015) demonstrated greater jaw variability (higher STI) as a function of word length (mono-, bi-, and trisyllabic: “pop,” “puppet,” and “puppypop,” respectively), while Terband et al. (2011) demonstrated greater variability of tongue tip movements in 6- to 9-year-old children with CAS (relative to TD peers). Furthermore, jaw deviances or instabilities (lateral movement range and variability) were found in the coronal plane, but not in the midsagittal plane for children with SSD or CAS relative to TD peers (Terband, van Zaalen, & Maassen, 2012). The findings of kinematic instability are in line with clinical observations (e.g., lateral jaw slide) in children with SSD and CAS (Namasivayam et al., 2013; Terband et al., 2012) and may be of diagnostic and therapeutic importance. In the following sections, we review perceptual, acoustic, and articulatory measures used to evaluate speech inconsistency in children with CAS.

Perceptual Measures

Background

To capture various types of error consistencies at the word and segmental level, several different formulas are reported in the literature (for details, please refer to Betz & Stoel-Gammon, 2005; Marquardt et al., 2004). For example, (in)consistency measured as a percentage of the total productions of a target word has been used by Dodd and colleagues (Dodd, 1995; Dodd et al., 2002) and Shriberg and colleagues (Shriberg, Aram, & Kwiatkowski, 1997a). This provides an index of “production consistency,” whereas the use of total error productions as the denominator is said to reflect “error consistency” (Betz & Stoel-Gammon, 2005; Iuzzini-Seigel, 2012). The numerator in such error consistency measures may also differ to capture (a) the proportion of errors, (b) consistency of error types, and (c) consistency of the most frequently used error type (Betz & Stoel-Gammon, 2005; Iuzzini-Seigel & Forrest, 2010; Shriberg et al., 1997a). The overall proportion of error productions

only provides a general impression of a child’s production accuracy and is not recommended as the only measure of consistency (Betz & Stoel-Gammon, 2005). In addition, the number of errors (e.g., number and variety of substitutions) and the most frequently used error type indicate the degree of variability in errors produced (in line with clinical impression of “inconsistent errors”; Betz & Stoel-Gammon, 2005).

Total Token Variability and Error Token Variability

Several procedures have been reported for assessing word-level inconsistency/variability, albeit with differing formulas and descriptions (Dodd, 1995; Ingram, 2002; Schumacher et al., 1986; Shriberg et al., 1997a; see Table 1). In a longitudinal study, Marquardt et al. (2004) assessed the accuracy, stability, total token variability (TTV), and error token variability (ETV) of whole-word productions in children with CAS (4;6–7;7) undergoing phonological treatment (for formula, see Table 1). Their study revealed that measures of stability and accuracy increased over time while variability (TTV) decreased. However, individual data showed clear session-to-session variability in patterns at the three time points for these children with CAS, with ETV emerging as the least consistent of the variables tested. The variability results obtained for children with CAS across time paralleled the results of single-word articulation testing and relational analysis of consonants and vowels in connected speech. For example, the child with higher levels of TTV and ETV and lower levels of accuracy and stability also had the lowest scores on relational analysis and articulation testing, possibly implying a relationship between severity of speech disorder and underlying speech motor variability (also see the ECI section).

With respect to validity, transcription-based word-level token-to-token consistency measures (e.g., TTV) were found to be moderately correlated with segmental-level (in) consistency assessments (e.g., Error Consistency Index [ECI]) but demonstrated low correlations with acoustic measures of phonetic variability (vowel formants, VOT, and coefficient of variation of word duration; Preston & Koenig, 2011). A comparison of interrater reliability suggests that broad phonetic transcriptions from spontaneous speech are more reliable than those of responses obtained from rapid picture-naming tasks (Marquardt et al., 2004; Preston & Koenig, 2011; see Table 1).

Token-to-Token Inconsistency Assessment: DEAP Inconsistency Subtest

Dodd and colleagues (Dodd et al., 2002; McIntosh & Dodd, 2008), as part of the DEAP Test, developed and standardized a 25-word picture-naming subtest to elicit word-level token-to-token inconsistency (see Table 2). In Token-to-Token Inconsistency assessment, a speaker is instructed to repeat the same utterance multiple times (three times) across a similar context, while their consistency of productions is scored as “same” (nonvariable) or “different” (variable). A production is considered variable if any of the productions differ in the three trials (Dodd et al., 2002). Dodd’s word-level

Table 1. Methodological details: total token variability and error token variability (Marquardt et al., 2004; Preston & Koenig, 2011).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | Six multisyllabic words (elephant, umbrella, strawberries, helicopter, thermometer, and spaghetti; Preston & Koenig, 2011) |
| (2) Tasks used to elicit those targets | Picture naming (Preston & Koenig, 2011) Spontaneously elicited connected speech samples using age-appropriate materials (Marquardt et al., 2004) |
| (3) Conditions in which responses are elicited | Quiet, with time pressure (rapid picture naming; Preston & Koenig, 2011) Quiet, no time pressure (Marquardt et al., 2004) |
| (4) The measures obtained from those responses | Total token variability: (number of variants – 1) / (number of tokens – 1) (Marquardt et al., 2004) Error token variability: (number of incorrect variants – 1) / (number of incorrect tokens – 1) (Marquardt et al., 2004) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No <i>Reliability:</i> Broad transcription reliability from spontaneous speech (10% of samples) = 86.22% (range: 75%–96.26%; Marquardt et al., 2004) Interrater reliability of total token variability scores based on phonetic transcription of rapid naming task with $r = .55$ (Preston & Koenig, 2011) |
| (7) Norm or reference data available? | No |

Token-to-Token Inconsistency assessment is a nominal measurement, and children with phonological disorders are classified as inconsistent or consistent, depending on whether or not they produced the same words consistently across three repetitions ($> 40\%$ = *inconsistent*). If inconsistency scores are greater than 40% (but see Iuzzini-Seigel, 2012, for higher cutoff $> 50\%$), along with the presence of other features, such as poor oromotor performance, poorer productions during imitation than spontaneous speech, consonant and vowel distortions, and atypical prosody, then a CAS diagnosis may be suspected (Dodd et al., 2002; see Table 2).

ECI

With respect to inconsistency measures at the segmental level, the ECI has been applied in a number of studies (Preston & Koenig, 2011; Tyler & Lewis, 2005; Tyler, Lewis, & Welch, 2003; see Table 3). The ECI is a raw score calculated as the sum of the total number of different error forms across all consonants and all word positions. A higher ECI score indicates a greater number of different error forms across a larger number of consonants, and a lower ECI score indicates fewer different error forms across a smaller number of consonants (Tyler & Lewis, 2005). The ECI measure is moderately–strongly correlated to token-to-token variability of repeated productions at word level and measures of speech severity, such as percent consonants correct (PCC; Preston & Koenig, 2011). Generally, correlation between PCC and ECI scores have been reported in the range of $r = -.58$ to $-.88$ in children with speech and language disorders (Tyler & Lewis, 2005; Tyler et al., 2003). Importantly, and as mentioned earlier (see the Error Inconsistency in CAS section), there are several studies that provide support for the notion that variability/consistency

measurements using such methods (e.g., ECI) may represent severity of the problem rather than disorder category (Betz & Stoel-Gammon, 2005; Forrest, Dinnsen, & Elbert, 1997; Forrest, Elbert, & Dinnsen, 2000; Tyler et al., 2006). With regard to reliability and validity, ECI score calculation has a high degree of reliability (99%; Tyler et al., 2003) and possibly addresses the same construct as other measures of speech severity (e.g., PCC; Tyler & Lewis, 2005; see Table 3).

TTR of Consonant Substitutions

TTR analysis is a measure of the number of types of productions to the total number of tokens produced (see Table 4). It indicates the number of different ways (i.e., inconsistency) a target form is produced by the child. Two variations of TTR analysis have been applied in both diagnostic and therapeutic contexts in the SSD and CAS populations. The segmental-level TTR measure, called *CSIP*, calculates a percentage based on the number of different error substitutes across all targets divided by the total number of erred productions across the whole inventory (Forrest & Iuzzini-Seigel, 2008; Iuzzini-Seigel, 2012). The *ISP* (Iuzzini-Seigel & Forrest, 2010) is derived from *CSIP* by modifying the denominator (of *CSIP*) from the total number of erred productions to the number of target opportunities. Validity of the *CSIP/ISP* measure has been demonstrated in few studies. Segmental-level *ISP* measure is correlated with the broader lexical-level word inconsistency scores ($r > .70$; Iuzzini-Seigel, 2012), which demonstrates construct validity. Interrater percent agreement scores for narrow transcriptions, as used in TTR analysis, is reported to be $> 90\%$ (Heisler, Goffman, & Younger, 2010; Iuzzini-Seigel, 2012; see Table 4).

Table 2. Methodological details: Word Inconsistency Assessment (Dodd et al., 2009).

Materials and methods

| | |
|--|--|
| (1) Stimuli or targets being analyzed | 25 words (ranging from one to four syllables) |
| (2) Tasks used to elicit those targets | Picture naming |
| (3) Conditions in which responses are elicited | Quiet, no time pressure, production of each target word in three separate trials, each trial separated by an intervening task (subsection of oral motor screen) or a short break (5 min) with conversation |
| (4) The measures obtained from those responses | Percentage of target words produced differently (word inconsistency score) |

Scientific basis

| | |
|---|---|
| (5) Standardized measurement protocol? | Yes |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Not specified in the DEAP test manual <i>Reliability:</i> Percent interrater agreement for Word Inconsistency Assessment based on whole-word narrow transcriptions from video/audio recordings was 91.64% ($SD = 5.76\%$; luzzini-Seigel, 2012) |
| (7) Norm or reference data available? | Reference data: $n > 40\%$ = inconsistent phonological disorder (Dodd, 2005; Tyler & Lewis, 2005) |

Note. DEAP = Diagnostic Evaluation of Articulation and Phonology.

Acoustic Measures

Acoustic Spatiotemporal Variability Indices

Assessment of speech variability via audio signals is clinically feasible even in difficult-to-test populations and has been recently proposed by several researchers (Anderson, Lowit, & Howell, 2008; Cummins, Lowit, & van Brenk, 2014; Howell et al., 2009; see Table 5). The acoustic STI is calculated in a similar manner to its kinematic variant but from the amplitude envelope derived from rectified and low-pass filtered speech audio recordings (Howell et al., 2009). As the source signal for variability calculation is the amplitude envelope, Howell et al. (2009) refer to this as *E-STI*. The *E-STI* measure captures the joint spatial and temporal variation in the patterning of speech amplitude envelopes over repeated utterances. For the *E-STI*, the sum of 50 *SDs* at 2% intervals is calculated over time- and amplitude-normalized repeated acoustic amplitude envelopes. While kinematic STI derived from single articulatory movement trajectories (or, in some cases, derived from interarticulatory distance measures) represent stability of underlying movement templates (Kleinow & Smith, 2000), the *E-STI* represents the summed output of respiratory, laryngeal, and articulatory subsystems. Lower *E-STI* values suggest less variability, a more robust and efficient speech subsystem coordination (Anderson et al., 2008; Cummins et al., 2014; Howell et al., 2009).

There is preliminary data to suggest that *E-STI* and kinematic STI are positively correlated and that *E-STI* is useful to discriminate speakers based on age and speakers who stutter from those who do not (Howell et al., 2009). A further methodological advancement over the STI/*E-STI* has been the nonlinear functional data analysis (FDA) procedure (Lucero, 2005; Lucero, Munhall, Gracco, & Ramsay, 1997; Ramsay & Silverman, 1997). The FDA procedure permits the estimation of spatial (or amplitude) and temporal

variability separately (Lucero, 2005). The FDA nonlinearly manipulates the time axis of acoustic (pitch, intensity, and formant tracks) or kinematic signals from successive utterances, such that their features are in alignment with each other. The amount of adjustment necessary to bring the signals into alignment provides an estimate of temporal variability, while the differences on the amplitude axis provide an estimate of spatial variability (Anderson et al., 2008; Howell, Anderson, & Lucero, 2010). Following time and amplitude alignment, temporal variability and spatial variability can be independently derived by averaging the standard deviation of the spatial and temporal errors across the signal (Anderson et al., 2008). Another recent development in the assessment of speech variability using acoustic recordings is the utterance-to-utterance variability (UUV) index (Cummins et al., 2014). For the UUV index, mel-frequency-scaled spectral coefficients are extracted from utterances, and a dynamic time-warping algorithm is used to map one utterance on to the other. The UUV index is a quantitative measure that represents the amount of warping (compression and stretching) required for the optimal mapping between the two utterances.

With regard to validity, *E-STI*, FDA, and UUV procedures have shown good comparability to other validated measures (e.g., kinematic STI) when investigating task demands on the speech motor system (e.g., changes in speech rate) and distinguishing type/severity of speech disorders (e.g., in dysarthria; Anderson et al., 2008; Mefferd, 2015; van Brenk & Lowit, 2012). These indices are also correlated with speech intelligibility ratings and standardized maximum performance tasks (e.g., diadochokinesis; Anderson et al., 2008; Cummins et al., 2014; Howell et al., 2010). Although these procedures have great potential for clinical use, they are yet to be applied to the CAS population. In terms of reliability, none of the studies examining

Table 3. Methodological details: Error Consistency Index (ECI; Preston & Koenig, 2011; Tyler & Lewis, 2005; Tyler et al., 2003).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | 64 words (included every English consonant at least twice—except /h/; Preston & Koenig, 2011) |
| (2) Tasks used to elicit those targets | Picture naming (Preston & Koenig, 2011) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure (Preston & Koenig, 2011) |
| (4) The measures obtained from those responses | ECI: Sum of all different error forms for all consonant phonemes combined (Preston & Koenig, 2011; Tyler & Lewis, 2005; Tyler et al., 2003) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Point -by-point consonant agreement = 87.3% (range: 81.5%–92.3%) Interrater reliability of ECI scores, $r = .98$ (Preston & Koenig, 2011) <i>Reliability:</i> Intra- and interreliability of error consistency scores derived from transcriptions = 99% (Tyler et al., 2003) |
| (7) Norm or reference data available? | Reference data: ECI range in preschool-age children with speech and language disorders: 12–70 ECI cutoff scores for children with speech and language disorders: variable group, upper quartile > 44.75; consistent group, lower quartile < 22.25 (Tyler & Lewis, 2005) |

these procedures reports any reliability scores related to segmentation of acoustic recordings or peak-picking algorithms (see Table 5).

VOT Variability

VOT is considered a robust and reliable acoustic temporal cue for distinguishing between voiced and voiceless plosive cognates (Auzou et al., 2000; Lisker & Abramson, 1964; see Table 6). It is defined as the time (in milliseconds) between the release of oral closure for plosive production and the onset of voicing (Lisker & Abramson, 1964) and reflects coarticulatory timing control between laryngeal and supralaryngeal mechanisms in speech production (Auzou et al., 2000; Whiteside et al., 2003). VOT and VOT variability have been investigated in children with SSDs arising from articulation and phonological impairments (Lundeborg, Nordin, Zeipel-Stjerna, & McAllister, 2015), speech motor issues (Yu et al., 2014), and apraxia of speech (AOS; Iuzzini-Seigel, Hogan, Guarino, et al., 2015).

Variability of VOT productions is usually calculated as the coefficient of variance of repeated productions. A few studies have used measures of VOT and VOT variability in the assessment of children with CAS. Compared to children with speech delay, children with CAS have been shown to produce shorter VOTs for voiceless stops, indicating a delay in acquisition of the voicing contrast (Iuzzini-Seigel, 2012; Iuzzini-Seigel, Hogan, Guarino, et al., 2015). As of yet, outcome measures related to VOT, such as absolute VOT length, VOT variability, or strength of voiced–voiceless contrasts, have not been correlated reliably to other outcome measures, such as intelligibility obtained with children with CAS.

With respect to reliability, one has to consider that VOT is a measurement of overlapping physiological events represented by strict, sometimes arbitrarily defined boundaries. As such, discrepancies in measurements within and across studies might be expected to some degree (Abramson

& Whalen, 2017). However, most studies report outcome measures obtained with high reliability (Iuzzini-Seigel, Hogan, Rong, et al., 2015; Lundeborg et al., 2015; see Table 6).

Articulatory Measures

Background on Kinematic Variability

The source or nature of articulatory variability depends on one's theoretical perspective. The motor control literature suggests that fluctuations of a value over repeated measurements (variability; Chau, Young, & Redekop, 2005) is an indicator of imprecise movements often associated with pathophysiology or an immature neuromotor system (e.g., A. Smith & Zelaznik, 2004). In theories such as the dynamical systems theory, variability also serves as an indicator of adaptability and flexibility in the system (Thelen & Smith, 1994; van Lieshout & Namasivayam, 2010). However, variability as a positive aspect of production has not really taken off in the field of SSD and CAS.

Objectively, movement variability has been described in the CAS literature in terms of discrete temporal or spatial parameters as related to single articulatory movements (e.g., standard deviations or covariance measures related to peak velocities, amplitudes, and duration of movements) and as measures of articulatory coordination (e.g., Grigos, 2009; Grigos & Patel, 2007; Nijland, Maassen, Hulstijn, & Peters, 2004; Terband et al., 2011, 2012). More recently, speech motor performance measures based on complete movement trajectories (from single articulators), called the *kinematic STI* (Kleinow & Smith, 2000), have been utilized. Researchers have also started to examine speech motor system (in)stability at the level of movement coordination within and between functional synergies. The specifics of these outcome measures are described in the subsections below.

Typically, optical (i.e., camera based using visible or infrared light) or electromagnetic articulography (EMA) systems have been used in children for tracking orofacial movements related to speech (Moss & Grigos, 2012; Terband

Table 4. Methodological details: type–token ratio: consonant substitute inconsistency percentage (CSIP)/inconsistency severity percentage (ISP; luzzini-Seigel, 2012; luzzini-Seigel & Forrest, 2010).

Materials and methods

| | |
|--|--|
| (1) Stimuli or targets being analyzed | 200–240 word probe list that provides 340–440 opportunities to produce all of the American English consonants in all naturally occurring word positions (luzzini-Seigel, 2012; luzzini-Seigel & Forrest, 2010) Stimuli also derived from the Goldman-Fristoe Test of Articulation 2 (GFTA-2) and the first trial of Word Inconsistency Assessment (Dodd et al., 2009) |
| (2) Tasks used to elicit those targets | Picture-naming task (if child is unable, then semantic cue or delayed imitation is carried out) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure |
| (4) The measures obtained from those responses | CSIP: percentage based on the number of different error substitutes across all targets divided by the total number of erred productions across the whole inventory (luzzini-Seigel, 2012; luzzini-Seigel & Forrest, 2010) ISP: percentage based on the number of different error substitutes across all targets divided by total number of productions (luzzini-Seigel, 2012; luzzini-Seigel & Forrest, 2010) |

Scientific basis

| | |
|---|---|
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Construct validity: high correlation between ISP ($r > .70$) and lexical-level word inconsistency scores (luzzini-Seigel, 2012) <i>Reliability:</i> Interrater percent agreement for narrow transcription $> 90\%$ (Heisler et al., 2010; luzzini-Seigel, 2012) |
| (7) Norm or reference data available? | Reference data: ISP score cutoff for CAS $> 17\%$ (luzzini-Seigel, 2012) |

Note. CAS = childhood apraxia of speech.

et al., 2011). Optical motion capture systems utilize small reflective markers (approximately 3 mm) that are placed on the child’s upper and lower lips, right/left/mid jaw, and lip corners to track speech-related movements. Other markers are placed on the forehead and nasion, which are used as reference to correct for head rotation/movements. An alternative to optical motion capture system is EMA. In EMA, the position and motion of sensor coils attached to speech articulators are tracked within a magnetic field. The sensor coils, typically around $4 \times 4 \times 3$ mm in size, are usually glued on the bridge of the nose, the maxillary gum ridge on the upper and lower lips, the mandibular gum ridge, and two or three points on the tongue. As the sensor coils are wired and directly glued on the articulators, this methodology is relatively invasive and might not be tolerated well by young children or infants. In comparison, the passive reflective markers used with optical motion tracking systems are unobtrusive, light, and well tolerated by young children and offer a more relaxed and naturalistic setting for data collection, especially in children. The limitation of optical motion capture systems is that they require a direct line of sight between the camera and the reflective marker and hence are only suited for the measurement of externally visible structures such as the jaw and lips. The operational principles of the optical motion capture and EMA systems have been elaborated elsewhere and are beyond the scope of this review (e.g., see Feng & Max, 2014; Yunusova, Green, & Mefferd, 2009).

Kinematic Spatiotemporal Variability Indices

For the STI, a sum of 50 *SDs* at 2% intervals is calculated over amplitude- and time-normalized repeated

movement trajectories (e.g., of the jaw or the lower lip) or individual movement cycles (cyclic STI; van Lieshout & Moussa, 2000; see Table 7). A lower STI value represents less variability, suggesting a robust and well-learned underlying movement template (Kleinow & Smith, 2000). With regard to stimuli and elicitation procedures, camera-based motion tracking of speech articulators in children has been limited to visible structures such as the jaw and lips and to words that comprise of bilabial consonants (e.g., pop, puppet, and puppypop; Moss & Grigos, 2012; buy bobby a puppy; A. Smith & Goffman, 1998). Stimuli with bilabial productions are also chosen with EMA systems for easier segmentation of position data (Terband et al., 2011). To acquire adequate data for measurement of articulatory variability (e.g., STI/cyclic STI), about 10–15 productions of the target stimuli are elicited. Most speech kinematic studies in children have elicited productions using picture naming, cloze sentence procedure (within a story retell game), or by direct/immediate word/sentence imitation tasks with auditory models (Grigos et al., 2015; Moss & Grigos, 2012; Sadagopan & Smith, 2008; Terband et al., 2011; see Table 7).

Covariance Measures

Moss and Grigos (2012) examined spatial coupling (calculated as absolute peak correlation coefficient [PC] between articulator pairs; i.e., between jaw and lower lip [J–LL], jaw and upper lip [J–UL], and upper and lower lip [UL–LL]) and temporal coupling (time required for peak spatial coupling; i.e., lag) as a function of word length (e.g., “pop,” “puppet,” and “puppypop”; see Table 8). A pair of articulators with a high degree of spatial and temporal coordination would yield high correlation coefficients

Table 5. Methodological details: acoustic spatiotemporal variability indices (Anderson et al., 2008; Cummins et al., 2014; Howell et al., 2009; van Brenk & Lowit, 2012).

Materials and methods

| | |
|--|---|
| (1) Stimuli or targets being analyzed | 20–25 repetitions of a phrase of which typically 10 are used for analysis: “Buy Bobby a puppy” (E-STI; Howell et al., 2009); “Well we’ll will them” (FDA; Anderson et al., 2008); “Tony knew you were lying in bed” (FDA/UUV; Cummins et al., 2014) |
| (2) Tasks used to elicit those targets | Phrase repetition |
| (3) Conditions in which responses are elicited | Quiet, self-selected comfortable/habitual speaking rate, twice as fast or half as fast as habitual speaking rate |
| (4) The measures obtained from those responses | Independent or combined temporal and spatial variability (E-STI/FDA/UUV) from audio recordings |

Scientific basis

| | |
|---|--|
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Results comparable to kinematic STI and negatively correlated with speech intelligibility ratings (Cummins et al., 2014; van Brenk & Lowit, 2012) <i>Reliability:</i> No |
| (7) Norm or reference data available? | No |

Note. E-STI = envelope-based spatiotemporal index; FDA = functional data analysis; UUV = utterance-to-utterance variability.

and low lag values. Moss and Grigos analyzed these measures in 3- to 6-year-old TD children and those with CAS and speech delay ($n = 6$ per group). There was no effect of group or Group \times Word interactions for PC and lag. Green, Moore, Higashikawa, and Steeve (2000) analyzed PC and lag in 1-, 2-, and 6-year-old TD children and adults. In general, 1- and 2-year-old children demonstrated greater spatial coupling between the UL–LL than between the lips and jaw pairs. The PC values indexing lip and jaw coupling (J–UL, J–LL) for 1-year-old children were very low, indicating weak coupling (values centered near zero). Spatial coupling values increased with age. With regard to lag-to-peak coefficient values, all articulatory movements (across pairs of articulators) were tightly coupled with mean lag values not > 29 ms for any age group (see Table 8).

Coefficient of Variation of Spatial and Temporal Coupling

Coefficient of variation of the PC (PCcov) and lag values (Lcov) from the Covariance Measures section were analyzed by Moss and Grigos (2012) for the following articulatory pairs: J–LL, J–UL, and UL–LL in 3- to 6-year-old TD children, those with speech delay, and children diagnosed with CAS ($n = 6$ per group; see Table 9). Significant main effects for group were found for PCcov and Lcov. The CAS group had significantly higher average PCcov and Lcov across utterances for J–LL coupling than the speech delay group (see Table 9).

Lengthened and Disrupted Coarticulatory Transitions Between Sounds and Syllables

Background

Coarticulation

Coarticulation refers to the phenomenon that the specific properties of articulatory movements are context

dependent as articulatory movements overlap in time and interact with one another. Acoustically, this manifests itself as the realizations of consecutive speech segments affecting each other mutually. The effect is bidirectional. Influences of a segment on a following segment are called *perseveratory* or *carryover coarticulation*, and influences of an upcoming segment on a preceding segment are known as *anticipatory coarticulation*. Furthermore, coarticulation is not limited to adjacent segments and can occur across syllables.

Coarticulation is the consequence of the inertia of the articulatory organs caused by their biomechanical characteristics and an economy of effort in articulatory planning influenced by biomechanical constraints (e.g., Recasens, 2004; Recasens, Pallarès, & Fontdevila, 1997), prosodic conditions (Cho, 2004; De Jong, 1995; Edwards, Beckman, & Fletcher, 1991), and syllable structure (e.g., Modarresi, Sussman, Lindblom, & Burlingame, 2004; Nittrouer, Munhall, Kelso, Tuller, & Harris, 1988; Sussman, Bessell, Dalston, & Majors, 1997). Furthermore, the amount of coarticulation depends on lexical frequency and, relatedly, the specific demands of the communication task (e.g., Farnetani & Recasens, 1997; Kühnert & Nolan, 1999). Perseveratory coarticulation has been found to reflect predominantly biomechanical constraints, whereas anticipatory coarticulation mainly reflects higher level phonetic processing (e.g., Daniloff & Hammarberg, 1973; Hertrich & Ackermann, 1995, 1999; Kent & Minifie, 1977; Whalen, 1990). Comparisons between carryover and anticipatory coarticulation effects are highly complicated, as both effects co-occur at multiple levels at approximately the same time. Moreover, the specific biomechanical constraints and syllabic position of the speech sounds involved play a role that is not straightforward and appears to be language specific, that is, some studies report stronger perseveratory as compared to anticipatory coarticulation whereas other studies report opposite effects (Beddor, Harnsberger, & Lindemann, 2002; Graetzer, Fletcher, & Hajek, 2015;

Table 6. Methodological details: voice onset time (VOT) variability (luzzini-Seigel, Hogan, Rong, et al., 2015; Whiteside et al., 2003; Yu et al., 2014).

Materials and methods

| | |
|--|--|
| (1) Stimuli or targets being analyzed | Five repetitions of CVC pseudowords (pVb), which sampled corner vowels (e.g., /pib/, /pub/; luzzini-Seigel, Hogan, Guarino, et al., 2015) 115 Repetitions of monosyllabic /pa/ (Yu et al., 2014) Five repetitions of 12 CVC target words with plosive consonants in syllable initial position (e.g., <i>pea</i> , <i>bee</i> , <i>tea</i> ; Whiteside et al., 2003) |
| (2) Tasks used to elicit those targets | Three repetitions of six minimal pairs (e.g., <i>pil–bil</i> , <i>tennis–dennis</i> ; Lundeborg et al., 2015) Imitation of recorded speech sample (luzzini-Seigel, Hogan, Guarino, et al., 2015) Cued (white circle on monitor) repetition task (Yu et al., 2014) Picture naming (luzzini-Seigel, Hogan, Guarino, et al., 2015) In carrier phrase “say ___ now” (luzzini-Seigel, Hogan, Guarino, et al., 2015) or “say ___ again” (Whiteside et al., 2003) |
| (3) Conditions in which responses are elicited | Quiet room, no time pressure |
| (4) The measures obtained from those responses | Duration in milliseconds of VOT measured described in terms of mean, SD median, median difference scores for voiced–voiceless cognates, COV, and skewness (luzzini-Seigel, Hogan, Guarino, et al., 2015) |

Scientific basis

| | |
|---|--|
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No <i>Reliability:</i> Intrarater reliability: ICC = .98–.99 (absolute error = 2.0–4.3 ms; luzzini-Seigel, Hogan, Guarino, et al., 2015); Cronbach’s alpha = .97 (Lundeborg et al., 2015). Interrater reliability: Pearson r = .97 (Whiteside et al., 2003); mean difference between raters = 17.19 ms (SD = 6.89 ms), Pearson r = .93 (Yu et al., 2014) |
| (7) Norm or reference data available? | Reference data: Mean COV values (in %) for voiced plosives approximately 20%–30% for typically developing children between 5;8 and 13;2 (years;months). Mean COV values (in %) for voiceless plosives approximately 15%–25% for typically developing children between 5;8 and 13;2 (Whiteside et al., 2003) <i>Typically developing 5-year-olds:</i> Mean COVs of 74% for /b/ and 51% for /d/. Mean COVs of 42% for /p/ and 34% for /t/ <i>3- to 5-year-old children with CAS:</i> Mean (SD) of COV = 56% (29) for /p/ and 52% (28) for /t/ <i>3- to 5-year-old children with phonological delay:</i> Mean (SD) of COV = 38% (19) for /p/ and 42% (25) for /t/ (luzzini-Seigel, 2012) |

Note. COV = coefficients of variation; CAS = childhood apraxia of speech; ICC = intraclass correlation coefficient.

Modarresi et al., 2004; Recasens & Pallarès, 2001; Sharf & Ohde, 1981).

Typical Development of Coarticulation

In typical development, coarticulatory patterns change as children become more adultlike in their speech production and improve spatiotemporal control. However, precisely how coarticulation changes during development has proved to be rather complex. Studies agree on the fact that coarticulation is more variable in the speech of children as compared to adults, but some studies report stronger coarticulation in children while other studies report that children exhibit less coarticulation than adults. At first glance, these results appear to be conflicting, but studies differ in experimental methodologies, procedures, language, stimuli, and age of participants. When examined closely, the results show a pattern in which “coarticulation that reflects poor temporal control or poor differentiation of structures decreases, whereas coarticulation that reflects language-specific efficiency increases” (ASHA, 2007, p. 8). More specifically, coarticulation decreases in general, as coordinative structures/functional motor synergies develop (e.g., Barbier et al., 2013; Noiray, Abakarova, Rubertus, Krüger, & Tiede, 2018; Noiray, Ménard, & Iskarous, 2013;

Sussman, Minifie, Buder, Stoel-Gammon, & Smith, 1996; Zharkova, Hewlett, & Hardcastle, 2011, 2012) and children move from a more global to a more segmental planning (Katz & Bharadwaj, 2001; Nijland et al., 2002; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Noiray et al., 2018; Siren & Wilcox, 1995). However, coarticulation increases (relatively) in certain contexts that are language specific, that is, depending on, for example, the phonological and articulatory specification of the segments involved (e.g., underspecified vowels exhibit more coarticulation; Nijland et al., 2002), prosodic patterns (e.g., stressed vowels exhibit less coarticulation; Nijland et al., 2002), and morphological structure or lexical frequency (e.g., higher frequent utterances show more coarticulation in adults but not in children; Song, Demuth, Evans, & Shattuck-Hufnagel, 2013). Furthermore, differences between anticipatory and perseveratory coarticulation in their developmental trajectories seem likely due to their differences in etiology, but the development of anticipatory and perseveratory coarticulation have not yet been compared directly in a single experimental design. In fact, little is known about the development of perseveratory coarticulation in general with the vast majority of studies focusing on anticipatory coarticulation (but see Song et al., 2013).

Table 7. Methodological details: spatiotemporal index (STI)/cyclic STI (cSTI; Grigos, 2009; A. Smith, Goffman, Zelaznik, Ying, & McGillem, 1995; Van Lieshout & Moussa, 2000).

Materials and methods

| | |
|--|---|
| (1) Stimuli or targets being analyzed | Eight to 15 productions of /papa/ and /baba/ produced with equal stress (Grigos, 2009) 10–15 productions of “pop,” “puppet,” and “puppypop” (Grigos et al., 2015; Moss & Grigos, 2012) Dutch words /pa:s/ and /spa:/ repeated for 5–12 s (three to six movement cycles per trial; Terband et al., 2011) |
| (2) Tasks used to elicit those targets | Object naming (Grigos, 2009) Closed-sentence procedure or respond to a “who”-question cued by a picture probe (Grigos et al., 2015; Moss & Grigos, 2012) Reiterated speech task–auditory model provided as needed (Terband et al., 2011) |
| (3) Conditions in which responses are elicited | No time pressure, play scenario (Grigos, 2009) Naturalistic productions embedded in a story retell game (Grigos et al., 2015; Moss & Grigos, 2012) |
| (4) The measures obtained from those responses | Syllable repeated at self-chosen normal, comfortable pace (Terband et al., 2011) Jaw, lower lip, and upper lip displacement trajectories (Grigos, 2009; Grigos et al., 2015) Lip aperture STI and lower lip–jaw STI (Moss & Grigos, 2012) cSTI for tongue tip, lower lip, and jaw (Terband et al., 2011) |

Scientific basis

| | |
|---|--|
| (5) Standardized measurement protocol? | No Segmentation based on zero crossing of jaw velocity trace (Grigos, 2009) Movement cycles (peaks/valleys in the position and velocity signals) were identified by automated algorithm using relative amplitude (10% of maximum amplitude) and time (a minimum interval of 0.5 s between successive events) criteria. Errors in automated peak/valley assignment were corrected manually (Terband et al., 2011) |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | Reference data: lower lip STI data on typically developing children and young adults for “buy bobby a puppy” phrase: $M(SD) = 24.1(4)$ for 4-year-old children, 18.5(5.7) for 7-year-old children, 13.6(2.5) for 20- to 27-year-old young adults (A. Smith & Goffman, 1998) |

In summary, the literature indicates that development does not involve a global increase or decrease in coarticulation. Speech motor development rather moves toward “flexible patterns of coarticulation” (Noiray et al., 2018, p. 1363; see also Noiray, Wieling, Abakarova, Rubertus, & Tiede, in press), which can differ depending on the phonetic and linguistic context. The point we want to make here, therefore, is that one should deliberate what the possible different outcomes would signify when assessing coarticulation, that is, would more or less coarticulation in a specific case indicate impaired, delayed, or more adultlike speech motor planning and programming?

Coarticulation in Children With CAS

As formulated in the CAS Technical Report, the speech of children with CAS is characterized by “lengthened and disrupted coarticulatory transitions between sounds and syllables” (ASHA, 2007, p. 4). First and foremost, children with CAS show coarticulation patterns that are not consistent, not typically immature, and highly idiosyncratic. Coarticulation effects usually change the characteristics of a speech sound in the direction of the neighboring speech sound. For 5- to 7-year-old children with CAS, however, coarticulation has been found to be both stronger and more extended, as well as the opposite, more segmental (or hyperarticulation), as compared to their TD peers (Maas & Mailend, 2017; Maassen, Nijland, & Van der Meulen, 2001; Nijland et al., 2002; Nijland, Maassen,

Van der Meulen, Gabreëls, et al., 2003; Sussman, Marquardt, & Doyle, 2000).

One factor that could be held responsible for this paradox is reduced phonological distinctiveness. The less distinctly speech sounds are produced, the weaker their possible coarticulatory influence on surrounding speech sounds. Children with CAS demonstrated weaker coarticulation in studies where they also showed a decreased differentiation of speech sounds as compared to their TD peers (stop consonants [Sussman et al., 2000] and vowels [Nijland et al., 2002; Nijland, Maassen, & Van der Meulen, 2003]). It is unclear why these studies found a decreased differentiation of speech sounds as not all studies do. Possibly, the decreased distinctiveness actually reflects coarticulatory effects in the opposite direction. In studies that feature similar phonological distinctiveness in the speech of children with CAS in comparison with TD children, coarticulation was found to be stronger and more extended (Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003). In a recent study, Terband (2017) investigated anticipatory coarticulation in [ə] as context-dependent F2 ratio relative to size of the produced phonetic contrast in the data set that was collected previously as part of the studies by Nijland and colleagues (Nijland et al., 2002; Nijland, Maassen, & Van der Meulen, 2003), thus taking the potential coarticulatory influence of the following speech sounds into account. The results showed increased coarticulation in the group of children with CAS ($n = 16$)

Table 8. Methodological details: covariance measures (Green et al., 2000; Grigos et al., 2015; Moss & Grigos, 2012).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | One-, two-, and three-syllable words (“pop,” “puppet,” and “puppypop”) repeated 10–15 times in random order (Moss & Grigos, 2012) |
| (2) Tasks used to elicit those targets | “Baba,” “papa,” and “mama” in 15 repetitions pseudorandom order (Green et al., 2000) Closed-sentence procedure or respond to a “who”-question cued by a picture probe (Moss & Grigos, 2012) |
| (3) Conditions in which responses are elicited | Reading for older children and imitation for younger children (Green et al., 2000) |
| (4) The measures obtained from those responses | No time pressure, naturalistic productions embedded in a story retell game (Grigos et al., 2015; Moss & Grigos, 2012) Peak correlation coefficient (PC) between articulator pairs and lag (time required for peak spatial coupling; Green et al., 2000; Moss & Grigos, 2012) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No Cross-correlation functions computed on the displacement traces |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No <i>Reliability:</i> 10% of data set was reanalyzed by the same experimenter for three coordinative indices (i.e., contribution to oral closure, coefficient, and lag). The mean absolute difference between first and second measurements of coefficient and lag was 0.012 and 3 ms, respectively. Pearson correlations between the first and second measurements ranged from 0.96 to 0.99. These findings suggest that the difference between the two measurements was negligible (i.e., good reliability; Green et al., 2000) |
| (7) Norm or reference data available? | Reference data: Mean (<i>SD</i>) of PC values and lag data from 3- to 6-year-old typically developing children for “puppypop” phrase: J–LL: PC: 0.62 (0.13), lag: 18.87 (2.77); J–UL: PC: 0.46 (0.08), lag: 27.86 (3.04); UL–LL: PC: 0.53 (0.06), lag: 26.78 (1.38; Moss & Grigos, 2012) Typically developing children (only data for 2- and 6-year-old typically developing children provided below due to space limitations; exact raw data unavailable; ~ = approximate values): J–LL: PC: ~0.3 to ~0.7, lag: ~ –0.02 to ~ –01; J–UL: PC: ~0.2 to ~0.4, lag: ~ –02; UL–LL: PC: ~0.6, lag: ~ –02 to ~ –01 (Green et al., 2000) Note: PC values close to one indicate a high degree of spatial coupling, while lag values close to zero indicate high levels of temporal coupling |

Note. J = jaw; LL = lower lip; UL = upper lip.

compared to TD children ($n = 8$), but this effect was limited to certain articulatory contexts. While TD children showed a differentiation in coarticulation between consonant contexts, the children with CAS did not. The results did not show any evidence of decreased coarticulation in CAS.

A second factor that is often put forward to explain the paradoxical findings is syllabic structure. The manipulation of syllable boundary or syllable shape revealed differences in the adjustment of the durational structure as a function of syllabic organization in children with CAS as compared to normally developing children (Maassen et al., 2001; Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003; see also Marquardt, Sussman, Snow, & Jacks, 2002). More specifically, the children with CAS did not show systematic durational adjustments to syllabic structure, and consistent intra- and intersyllabic temporal structures were missing (Maassen et al., 2001; Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003; see also Marquardt et al., 2002). However, the differential effects of syllable structure on coarticulation are less clear. Children with CAS did not show a significant coarticulation effect across syllable boundaries, while TD children showed stronger intersyllabic coarticulation as compared to adults. However, this lack of a group-level effect could very well be due to the

large variability in the children with CAS—both within groups and within subjects (Nijland et al., 2002). In direct comparison, no differences were found between inter- and intrasyllabic coarticulation, neither in the children with CAS nor in their TD peers (Maassen et al., 2001; Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003). Although these studies did not contain an adult control group, such an effect has been reported for adults in the literature (e.g., Modarresi et al., 2004; Nittrouer et al., 1988; Sussman et al., 1997). However, the location of syllable boundary did have an effect, and intersyllabic coarticulation was found to be stronger in V/CC (e.g., /zə sxi:t/; “ze schiet”) than in VC/C (e.g., /zəs xit/; “zus giet”) sequences for both groups of children (Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003). In summary, whereas syllabic structure has been found to have a different effect on temporal organization (the durations of the speech sounds) in 5- to 7-year-old children with CAS compared to their TD peers, it does not have a differential effect in terms of coarticulation.

Perceptual Measures

Identification of Gated Stimuli

Due to the transient nature of the acoustic signal, speech characteristics involving fine-grained phonetic detail

Table 9. Methodological details: coefficient of variation of spatial and temporal coupling (Moss & Grigos, 2012).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | One-, two-, and three-syllable words (“pop,” “puppet,” and “puppypop”) repeated 10–15 times in random order |
| (2) Tasks used to elicit those targets | Closed-sentence procedure or respond to a “who”-question cued by a picture probe (Moss & Grigos, 2012) |
| (3) Conditions in which responses are elicited | No time pressure, naturalistic productions embedded in a story retell game (Grigos et al., 2015; Moss & Grigos, 2012) |
| (4) The measures obtained from those responses | Coefficient of variation of peak correlation coefficient (PCcov) between articulator pairs and coefficient of variation for lag (time required for peak spatial coupling; Lcov; Moss & Grigos, 2012) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | Reference data: Mean (<i>SD</i>) of jaw–lower lip PCcov and Lcov data of 3- to 6-year-old typically developing (TD), CAS and children with speech delay for the phrase “puppypop”: TD: PCcov: 0.36 (0.15), Lcov: 0.65 (0.27); speech delay: PCcov: 0.25 (0.10), Lcov: 0.35 (0.14); CAS: PCcov: 0.54 (0.22), Lcov: 0.73 (0.30; Moss & Grigos, 2012) |

such as coarticulation are very difficult to assess perceptually (see Table 10). Ziegler and von Cramon (1985) used a vowel identification task in which a panel of nine trained listeners were presented with gated speech segments containing parts of increasing length of three test words with the form /gətV:tə/ with target vowels (/i, y, u/) and were asked of which test word the segment was the beginning of (see Table 10). The percentage of correct identification is indicative for the amount of coarticulatory information that is contained in the stimulus and can be analyzed as a function of stimulus length and compared between speakers with and without speech disorder. Examining the productions of a patient with AOS compared to three control speakers, Ziegler and von Cramon found that the onset of the vowel gesture was delayed in /i/ and /y/, whereas for /u/ the differences with the control speakers were not as pronounced. These results indicate a reduced anticipation of the upcoming articulatory movement (lip spread in case of /i/ and lip rounding in case of /y/) in the patient with AOS. Using a similar gating technique, Southwood, Dagenais, Sutphin, and Garcia (1997) replicated this finding of reduced anticipatory coarticulation in another apraxic patient.

This measure has not been used in children and only sparsely in populations with speech disorders in general. Its potential for use in clinical settings is limited as the procedure yields 90 stimuli per speaker and requires an elaborate perception experiment with a panel of trained listeners.

Acoustic Measures

Background

There is a large body of studies involving acoustic measurements of coarticulation, typically comparing specific spectral characteristics of the acoustic signal across different contexts. Measurements can focus on the acoustic spatial domain (how much the acoustics are influenced) or the temporal domain (how far the influence reaches).

Acoustic outcome measures to assess coarticulation are stimuli specific, and which measure is appropriate depends on the speech sounds that are involved. In vowels, coarticulation can be calculated with mean formant frequencies measured over a short time window (10–30 ms) at different parts of the speech sound, typically comprising onset, midpoint, and offset. While primarily formant frequencies at midpoint are indicative for realized vowel quality and articulatory positioning, other parts of the vowel can be used to investigate the range of the coarticulatory influence. Exact definitions of onset and offset vary between studies but are usually at about 20%–30% and 70%–80% of the vowel, respectively. Few studies have focused on sonorants and liquids, but coarticulation in these speech sounds can be measured similar to vowels. The same principle applies to fricatives, provided that the calculations are not based on formant analysis but on the spectral moment of the friction noise. When little spectral information is available, such as in the case of plosives, place of articulation should be derived from the formant trajectories in the consonant-to-vowel or vowel-to-consonant transition.

Acoustic measurements of coarticulation typically involve the first three formants, with F2 as the most prominent measure of interest. Under the assumption of an idealized vocal tract model, changes in vocal tract shapes during coarticulation might be obtained from tracing the formant contours over time. The most prominent relationships in the context of coarticulation are the following. First formant frequencies are inversely related to tongue height, that is, high vowels have low F1 values and low vowels have high F1 values. Second formant frequencies are related to tongue advancement, that is, front vowels have high F2 values and back vowels have low F2 values. Third formant frequencies have been found to be related to lip rounding in front vowels, with low F3 values present in rounded vowels and high F3 values present in unrounded vowels (Harrington, 2010). With respect to

Table 10. Methodological details: identification of gated speech stimuli (Ziegler & von Cramon, 1985).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | Six repetitions of three words /gətV:ts/ with target vowels (/i, y, u/); each of which five gating segments of increasing length were extracted |
| (2) Tasks used to elicit those targets | Imitation (model produced by experimenter) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in carrier phrase (“Ich habe /.../ gehört,” “I have heard /.../”) |
| (4) The measures obtained from those responses | Percentage /i, y, u/ responses per gating segment in an identification task by a panel of trained listeners |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | No |

voiced consonants, transitions of F2 have been found to be a relatively reliable indicator of place of articulation, with increasing F2 trajectories for labial consonants to decreasing F2 trajectories for dorsal consonants (e.g., Kewley-Port, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). As such, F2 has been found in general to be more sensitive to coarticulation than F1 and F3 (Öhman, 1966).

With regard to stimuli and elicitation procedures, many studies have used schwa–CV(C) sequences. When interested in consonant production, the unspecified, neutral vowel limits systematic carryover coarticulation and schwa proves to be very sensitive to anticipatory coarticulation, making it a very suitable object of study itself (Nijland et al., 2002; Nittrouer, 1993). Corner vowels are often included in the assessment materials, as they are most distinctive within the F1–F2 space. When studying vowel-to-vowel coarticulation, consonant context is important to consider as recent results have suggested that deviant coarticulation in children with CAS compared to TD children might be limited to certain articulatory contexts (Terband, 2017).

A further consideration is that measuring formants in children can be difficult due to their relatively high fundamental frequencies, which generate widely spaced harmonics, leading to an undersampling of the vocal tract transfer function, and may cause first and second formants to blend (Lee, Potamianos, & Narayanan, 1999; Nijland et al., 2002; Story & Bunton, 2016). This has been found to be particularly problematic in earlier studies using speech processing programs with limited linear predictive coding and visualization capabilities (Bennett, 1981; Bickley, 1986; Nittrouer et al., 1989). Solutions to this measurement problem, while becoming less urgent with modern speech processing software, are still researched, for example, by extracting the spectral envelope through improved spectral filtering techniques (Story & Bunton, 2016).

Children with CAS might display reduced articulatory rate and reduced size or amplitude of articulatory movements, which may complicate interpretations of coarticulatory effects: Both reduced articulation rate and reduced speech movements may contribute to the appearance of

reduced coarticulation. These factors require appropriate attention when designing and analyzing speech tasks employed to assess coarticulation in CAS (Hardcastle & Tjaden, 2008).

The three most prominent acoustic techniques to evaluate coarticulation are F2 ratios, first moment coefficients, and F2 locus equations. Since F2 ratios and first moment coefficients are usually reported side by side, these outcome measures will be discussed jointly, followed by a separate subsection on F2 locus equations.

F2 Ratios and First Moment Ratios

Coarticulation in children’s speech has mainly been quantified by using the center of gravity (also named *spectral centroid* or *first moment of the spectral distribution*) and fricative F2 frequencies as outcome measures (Nittrouer et al., 1989; see Table 11). Typically, stimuli with varying fricative spectral distributions and vowels with lip-spreading and lip-rounding features are used, for example, /sisi/, /ʃiʃi/, /susu/, and /ʃuʃu/. Coarticulation is usually quantified by calculating F2 ratios: dividing mean F2 values in /i/ utterances by mean F2 values in /u/ utterances averaged across a series of repetitions (see Table 11). The F2 ratios provide a measure to distinguish the utterances. High F2 ratios in the vowels indicate large distinctions between vowels, and the F2 ratios in the measurement points preceding the vowel reflect the coarticulation effect of the upcoming vowel (Nittrouer et al., 1989). It has been found, however, that centroids tend to be a relatively poor measure of fricative vowel coarticulation but are rather a measure of anticipatory lip rounding (Nittrouer et al., 1989; Soli, 1981).

Despite the fact that lengthened and disrupted coarticulatory transitions has been identified as one of the main criteria in CAS, the literature on coarticulation is, as of yet, relatively modest in size, compared to the literature investigating coarticulation in neurotypical children and adults (Hardcastle & Tjaden, 2008). A number of studies have used acoustic measures of coarticulation in the assessment of children with CAS. As of yet, no coherent picture can be drawn with respect to coarticulatory behavior in CAS. Compared to their TD peers, children with CAS have found to display earlier and stronger anticipatory

Table 11. Methodological details: first moment ratio/F2 ratio (Maas & Mailend, 2017; Nijland et al., 2002; Nittrouer et al., 1989).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | Eight repetitions of four reduplicated syllables (/CVCV/) consisting of a fricative (/s/, or /ʃ/) followed by a vowel context (/i/, or /u/; Nittrouer et al., 1989) Six repetitions of 12 /də'CV/ syllables consisting of an initial stop (/b/, /d/, /s/, and /x/) followed by three final vowel contexts (/i, a, u/; Nijland et al., 2002) Six repetitions of 12 /CVb/ syllables consisting of an initial fricative (/s, z, ʃ/) followed by three final vowel contexts (/i, a, u/; Maas & Mailend, 2017) |
| (2) Tasks used to elicit those targets | Imitation (model produced by experimenter) Accompanied by a picture (Nittrouer et al., 1989) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure Items in isolation (Nittrouer et al., 1989) Items in carrier phrase ("Hé /də'CV/ weer" [he...wl:r] ("hey...again"; Nijland et al., 2002) Items in carrier phrase ("It's the /CVb/ again"; Maas & Mailend, 2017) |
| (4) The measures obtained from those responses | Ratio of F2 frequencies in different vowel contexts (Nittrouer et al., 1989) Ratio of F2 frequencies in different vowel contexts at /ə/ midpoint, /ə/ end, C onset, CV transition onset, CV transition end, and V midpoint (Nijland et al., 2002) Ratio of first spectral moment (Maas & Mailend, 2017) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No <i>Reliability:</i> Interinvestigator differences in segmentation: 12.2 ms; correlation between segmentation markers: $r > .78$ (Nijland et al., 2002) Interinvestigator differences in segmentation: 1.2 ms (onset) and 1.4 ms (offset); correlation between segmentation markers: $r > .99$ (Maas & Mailend, 2017) Validity and reliability of F2 values by a postprocessing procedure of outlier removal (Nijland et al., 2002) |
| (7) Norm or reference data available? | Reference data: F2 frequencies, fricative ratios, and vowel context ratios for /si/, /ʃi/, /su/, and /ju/ are reported for eight participants per age group for adults (four males, four females) and 3-, 4-, 5-, and 7-year-old TD children (Nittrouer et al., 1989) Mean midpoints and width of ranges of F1 and F2 and variability of F2 of schwa and vowels for children with CAS, TD children, and adult females are reported (Nijland et al., 2002) F ratios and V ratios for /si/, /ʃi/, /su/, and /ju/ are reported for adults, TD children, and children with SSD (Maas & Mailend, 2017) |

Note. TD = typically developing; CAS = childhood apraxia of speech; SSD = speech sound disorder.

coarticulatory vowel effects during a preceding consonant (Maassen et al., 2001), display higher variability in the amount of coarticulation, and display reduced distinctions between different vowels (Nijland et al., 2002). Findings of reduced contrasts have been reproduced when studying fricative productions in children with SSD, independent of SSD subtype (Maas & Mailend, 2017). Abnormal (greater and reduced) coarticulation was observed only in children diagnosed with CAS (Maas & Mailend, 2017).

Locus Equation Metric

The locus equation metric was originally conceived by Lindblom (1963), as cited in Sussman, McCaffrey, and Matthews (1991), in the search for an invariant cue of place of articulation in stop consonants, independent of vowel context (Sussman et al., 1991; see Table 12). While initially based on voiced stops, it has been found to be an effective descriptor of place of articulation for consonants with other manners of articulation as well (Fowler, 1994; Sussman, 1994; Sussman & Shore, 1996; but see also Brancazio & Fowler, 1998) and has been shown to be stable across languages (Krull, 1988; Sussman, Hoemeke, & Ahmed, 1993). Furthermore, the measure has been shown to work in adults and in children as young as 1.5 years old (Chang, Ohde,

& Conture, 2002; Gibson & Ohde, 2007; Sussman, Hoemeke, & McCaffrey, 1992; Sussman et al., 1996).

Locus equations are based on the correlation between the values of F2 at vowel onset and vowel midpoint in CV sequences for a given consonant across vowel contexts. Lindblom (1963) found that the relationship between F2 at onset and F2 midvowel can be described by a linear regression equation: $F2_{onset} = k \times F2_{vowel\ midpoint} + c$, where k is the slope of the regression line and c is the y intercept (the value where the regression line crosses the y -axis at $x = 0$; Lindblom, 1963, as cited in Sussman et al., 1991). Regression slope and y intercept can then be used to quantify anticipatory coarticulation in CV utterances where a steeper slope (i.e., a larger value of k) and a lower y intercept (a smaller value of c) indicate more coarticulation (Krull, 1989). In general, regression slope and y -intercept values show a strong correlation. Alveolar and dental productions, for example, typically feature shallower slopes and higher y intercepts, while bilabials typically feature steeper slopes and lower y intercepts. Approximants, however, form an exception and typically feature slopes near zero with varying F2 onset loci exclusively described by varying y intercepts (Sussman, 1994; Sussman & Shore, 1996).

Although locus equations have only been used sparsely in children with CAS and children with speech disorders in general, they show great potential. Using locus equations, Sussman and colleagues demonstrated decreased differentiation of stop place of articulation as well as a pattern of decreased and less stable coarticulation across stop consonants in five children with CAS compared to children with typical development (Sussman et al., 2000), while Chang et al. demonstrated that children who stutter do not differ from their TD peers in terms of degree of coarticulation (Chang et al., 2002).

Reliable locus equations can be obtained using several tasks and stimuli. The elicitation method used to obtain responses appears to have little effect on locus equations (Chang et al., 2002; Gibson & Ohde, 2007; Sussman et al., 1992). The original study of Sussman et al. (1992) used an imitation task with the stimuli embedded in carrier phrase “It’s a /CVt/ again,” while Chang et al. (2002) successfully used a picture-naming task and Gibson and Ohde (2007) used spontaneous elicitation and imitation during free-play and child-centered activities with toys and pictures in their study with toddlers from 1.5 years old. While elicitation method is somewhat flexible, a requirement that is crucial is that the stimuli should contain enough variation in vowel context. It is not clear what constitutes the exact minimum number of vowels needed to reliably calculate locus equations. However, Nijland et al. (2002) reported that using only the three corner vowels /i, a, u/ did not result in reliable slope calculation. It is therefore advised to obtain minimally three repetitions of six dissimilar vowels, as described by Sussman et al. (see Table 12).

Articulatory Measures

Background

With respect to coarticulation, articulatory analyses would have value for understanding CAS as a motor speech disorder but, to date, have not been applied in this population. A wide variety of techniques are available that have been used for tracking speech movements and articulatory positioning in children. Similar to articulatory

measures of inconsistency, techniques include EMA and optical motion capture systems. Since the technical background, general procedures, and methodological considerations regarding these techniques have been described in the Background on Kinematic Variability and Kinematic Spatiotemporal Variability Indices sections, we will only highlight additional aspects that are specific when studying coarticulation. In addition, electropalatography (EPG; Timmins, Hardcastle, McCann, Wood, & Wishart, 2008) and ultrasound imaging systems (e.g., Noiray et al., 2018; Song et al., 2013; Zharkova et al., 2011, 2012) have been used to assess coarticulation in children. EPG utilizes an individually tailor-made artificial palate, placed inside the mouth against the speaker’s hard palate, containing electrodes that record timing, location, and (in modern systems) pressure of lingual contact. As such, EPG can be used to measure spatiotemporal aspects of tongue–palate constrictions but does not track articulatory movements. Oppositely, ultrasound can be used to track tongue movements but is less suitable to visualize and quantify lingual constrictions. With ultrasound, a sonic transducer is placed head-mounted, tightly under the chin. The transducer emits high-frequency sound waves and records their echo as the sound waves are reflected by bodily fluids and soft tissue, such as the lingual musculature. Ultrasound is gaining popularity quickly due to its relatively low-cost and limited invasiveness. Although ultrasound records full-tongue contours, its time resolution is limited in comparison with EPG and EMA systems. Other imaging techniques include X-ray microbeam and magnetic resonance imaging, but these are generally not considered suitable for children. The operational principles of the EPG and ultrasound systems have been elaborated elsewhere and are beyond the scope of this review (e.g., Cleland, McCron, & Scobbie, 2013; Gibbon & Lee, 2007; Zharkova, 2013).

Articulatory measures of coarticulation basically comprise two approaches and focus either on articulatory timing or on articulatory positioning. Articulatory timing measures assess the temporal coordination between speech movements and operationalize coarticulation as the overlap in time between the realization of consecutive

Table 12. Methodological details: locus equation metric (Sussman et al., 1992).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | Three to six repetitions of 18 /CVt/ syllables consisting of an initial stop (/b/, /d/, and /g/) in six vowel contexts (/i/, /I/, /æ/, /a/, /ʌ/, and /u/) |
| (2) Tasks used to elicit those targets | Imitation (model produced by experimenter) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in carrier phrase (“It’s a /CVt/ again”; Sussman et al., 1992) |
| (4) The measures obtained from those responses | Regression slope and y intercept of the linear relationship between the frequencies F2 at onset and F2 at vowel midpoint |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No <i>Reliability:</i> Interinvestigator differences in F2 frequencies: 97.2 Hz; correlation between F2 measurements: $r > .95$ |
| (7) Norm or reference data available? | No |

articulatory movements. The amount of overlap can be calculated based on either offset and onset or midpoints of either movements or target configurations, depending on the sequence of speech sounds involved and the specific research question. In principle, these overlap measures require simultaneous tracking of movements or constrictions made by different articulators. Lingual coarticulation can therefore only be measured with EMA, ultrasound, or EPG. In addition to EMA, however, anticipatory or carry-over effects of lip rounding can be assessed using an optical motion capture system when combined with onset or offset timing of the coarticulatory context segment based on acoustics.

Measures of articulatory positioning, on the other hand, assess differences in realized place of articulation across different contexts, similar to the acoustic measures of formant ratios and locus equations described above. A strong advantage of articulatory over acoustic measures of coarticulation is that they are based on direct information of the positioning of the articulators. Since the articulatory positioning does not need to be derived from the acoustic signal, coarticulatory influences on movement targets and trajectories can be investigated even if they do not manifest themselves acoustically. Where early kinematic studies relied on visual inspection of movement trajectories across stimuli (e.g., Katz & Bharadwaj, 2001; Katz, Machetanz, Orth, & Schönle, 1990), quantitative measures have been developed in recent years, and a variety of different techniques have been used to this end. Coarticulatory context segments do not necessarily need to be assessed kinematically, and the quality of the realized segments could be verified perceptually or acoustically (e.g., Weismer, Yunusova, & Westbury, 2003).

Over the years, a large set of movement-based measures of coarticulation has been reported in the literature, often conceptually similar but adapted or adjusted based on technological progress in data collection and processing. In this tutorial, we will focus on the more recent versions of measures that can be used with the modern systems.

Temporal Gestural Overlap

The first studies to investigate coarticulation through measures of articulatory timing used EPG (e.g., Butcher, 1989; Butcher & Weiher, 1976; Hardcastle, 1985), later followed by EMA (Katz et al., 1990; see Table 13). These early studies relied on a comparison of the timing of articulatory movements in different contexts but did not quantify temporal overlap as such. Measures of overlap in time between speech movements were developed a decade later, driven by the need to normalize for durational differences (Gibbon, Hardcastle, & Nicolaidis, 1993) and as a part of the different, more general endeavor to investigate coordination of speech movements in the theoretical framework of articulatory phonology (inter- and intragesural coordination; e.g., Chitoran, Goldstein, & Byrd, 2002; Kühnert, Hoole, & Mooshammer, 2006).

Most studies investigating temporal overlap have focused on consonant–consonant sequences. Using EPG,

Gibbon et al. (1993) formulated the overlap index as the overlap between the articulatory constrictions of consecutive consonants relative to the duration of the first consonant. In formula: $(\text{approach closure C1} - \text{approach closure C2} / \text{approach closure C1} - \text{release closure C1}) \times 100$. Herein, *approach* and *release* are defined as the starting points of the first palatal contact and the release of full closure for the consonant constriction, respectively. A value of < 100 indicates that the articulation of the two segments overlap (the lower the value, the stronger the coarticulation), while a value of > 100 indicates a gap between the two segments. Using EMA, Kühnert et al. (2006) calculated the measure of overlap between articulatory constrictions with a different formula as $(\text{Offset C1} - \text{Onset C2} / \text{Offset C2} - \text{Onset C1}) \times 100$. Positive values indicate overlap between the two segments, while negative values indicate a lag. Onset and offset are defined as the start- and endpoints of the movement plateau, meaning the start and end of the full constrictions, and thus correspond to “approach” and “release” in the EPG formula. The important difference between these two ways of calculating the overlap/lag between two consecutive segments is that, in Gibbon and colleagues, lag is relative to the duration of C1, whereas lag is relative to the duration of C1 + C2 in the method by Kühnert and colleagues. We believe the latter is to be preferred since it is less sensitive to durational adjustments of individual segments. In principle, the same measure could be used for studying anticipatory coarticulation in CV or VC sequences by substituting C1 or C2 constriction offset and onset, with the offset and onset of the movement plateau of the vowel. However, precisely establishing the points of movement plateau or constriction offset and onset is much more difficult in vowels than in consonants. The technical details go beyond the scope of this tutorial, but it should be noted that validity and reliability of the measurement procedures have yet to be established in large sample studies.

Regarding stimuli, many studies have focused on initial /k/ sequences in real words. In principle, however, any heterorganic sequence of articulatory movements is possible (if EPG is used, with the obvious limitation that the speech sounds must involve lingual–palatal contact). Regarding task and elicitation procedures, studies with children have used real words preceded by an indefinite article repeated from a wordlist read by the experimenter at an habitual rate (Timmins et al., 2008). The target words were mixed with filler items, and the whole list was repeated 10 times (see Table 13). The recorded words were subject to a qualitative phonological analysis, and incorrect productions in which not all segments were realized were removed from the analysis.

Articulatory Positioning: Anticipatory Lip Rounding

Labial anticipatory coarticulation has mainly been investigated in adults in the context of theories to account for cross-linguistic differences in anticipatory rounding behavior but has also been successfully assessed in 3.5- to 8-year-old children (Noiray, Cathiard, Abry, & Ménard, 2010; Noiray,

Table 13. Methodological details: temporal gestural overlap/lag (Kühnert et al., 2006; Timmins et al., 2008).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | /CC.../ sequences comprising real words (e.g., “clock”) |
| (2) Tasks used to elicit those targets | Read at self-chosen, habitual, rate; 10 iterations of each target word |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in carrier phrase (“I see...”; Kühnert et al., 2006) or preceded by an article (“a...”; Timmins et al., 2008) |
| (4) The measures obtained from those responses | Gestural overlap/lag = $(t_{\text{offsetC1}} - t_{\text{onsetC2}} / t_{\text{offsetC2}} - t_{\text{onsetC1}}) \times 100$ (Kühnert et al., 2006) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | No |

Ménard, Cathiard, Abry, & Savariaux, 2004; see Table 14). Two parameters indicative for lip rounding have been investigated in this respect, lip protrusion and lip constriction, of which the latter has been consistently shown to be more reliable (Noiray et al., 2010; Noiray, Cathiard, Ménard, & Abry, 2011; Ménard, Cathiard, Dupont, & Tiede, 2013). Where earlier studies used a combination of three-dimensional optical (infrared light) and video recordings, later studies rely on video-based registration only (e.g., Ménard et al., 2013).

In this technique, lip constriction is measured as between-lips area based on the labial contours. Speakers’ lips are marked with a blue lipstick to maximize visual contrast, and a purpose-designed video analysis software automatically tracks and processes labial shapes. The time resolution depends on the camera, but the software doubles the frame rate of the camera (which means that, with modern ordinary equipment, rates of $\geq 60\text{Hz}$ are easily attainable). The operationalization of anticipatory coarticulation is strongly intertwined with the stimuli, consisting of $V1C_nV2$ sequences ($/i/C_n/y/$ or $/i/C_n/u/$), in which C_n varied from zero to three consonants. In these sequences, anticipatory vowel behavior is assessed through the relation between the total duration of the rounding gesture in the final vowel and the duration of the obstruction interval or, in other words, is measured by how early in the utterance lip rounding starts. The duration of the constriction gesture is based on the video data, with the onset marked by a

the point at which lip area shows a 10% decrease following the maximum area and offset by the point of a 10% increase following the minimum lip area. The duration of the obstruction interval is based on the acoustic signal with V1 offset and V2 offset determined from the spectrogram.

The stimuli used by Noiray et al. (2004, 2010) contain intervocalic consonant sequences of increasing length, which was specific to their study testing theoretical hypotheses about the temporal expansion of lip rounding as a function of intervocalic obstruction interval duration (see Table 14). For children with CAS, some of these complex intervocalic consonant clusters might be (too) difficult to produce. In principle, however, any intervocalic consonant sequence could be used, as long as the consonants are phonologically neutral with respect to rounding and the clusters are phonologically legal in the testing language.

Articulatory Positioning: Mean Distance Across Set/Context

A first type of measure of coarticulatory influences on articulatory positioning is the absolute distance between the position of an articulator during the production of a speech sound in different contexts and has been mainly used to investigate lingual coarticulation (see Table 15). Distance measures can be based on tongue contour as a whole or on

Table 14. Methodological details: anticipatory lip rounding (Noiray et al., 2004, 2010).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | 10–12 repetitions of $/iC_ny/$ sequences in which C_n varied from zero to three intervocalic consonants (in French forming the names [iy], [isy], [iky], [iksy], [ikry], [itkry], [iskry], and [iksty]) |
| (2) Tasks used to elicit those targets | Imitation, prompted by the experimenter |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items embedded in carrier sentences |
| (4) The measures obtained from those responses | Anticipatory lip rounding = total duration of the rounding gesture in the final vowel / duration of the obstruction interval |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | No |

Table 15. Methodological details: mean distance across set/context (Kim et al., 2018; Zharkova et al., 2011, 2012).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | 10 repetitions of CV syllables consisting of a fricative (/s/ or /ʃ/) followed by a vowel context (/i/, /a/, and /u/; Zharkova et al., 2011, 2012) Five repetitions of /ə/ preceded or followed by four CVC words consisting of two voiced alveolar plosives /d/ with the vowels (/i/, /u/, /a/ and /æ/; Kim et al., 2018) |
| (2) Tasks used to elicit those targets | Reading/picture naming (text + image on screen; Kim et al., 2018; Zharkova et al., 2011, 2012) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in carrier phrase (“It’s a...Pam”: Zharkova et al., 2011, 2012; “Get...a puppy” and “Put a...here”: Kim et al., 2018) |
| (4) The measures obtained from those responses | Mean across set/context distance (in millimeters) of tongue contours (Zharkova et al., 2011, 2012) or tongue body position (Kim et al., 2018) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | Reference data: mean across set/context distance of tongue contours for /si/, /su/, /sa/ and /ji/, /ju/, /ja/ are reported for 10 participants per age group for adults (gender not reported) and 6- to 10-year-old TD children (Zharkova et al., 2011, 2012) |

Note. CV = consonant–vowel; TD = typically developing.

specific parts of the tongue (e.g., flesh-point markers on tongue tip, body and dorsum [EMA], highest point in tongue body [ultrasound], center point of contact [EPG]). Using ultrasound tongue imaging, Zharkova et al. (2011, 2012) quantified coarticulation as the mean nearest neighbor distance between tongue curves at midpoint of the production of the initial fricatives /s/ and /ʃ/ in two vowel contexts, calculated as the Euclidean distance from each point in one curve to the nearest point in the second, comparison curve. Coarticulatory distance between single points instead of contours, such as EMA coil position or EPG center point of contact, could be calculated in the same way.

A similar but slightly different approach was used by Kim, Coalson, and Berry (2018) in investigating articulatory measures of anticipatory and carryover lingual coarticulation in (/ə)CVC(/ə) sequences with EMA (see Table 15). Instead of comparing tongue position in two contexts, they compared each /ə/ production with the speaker-specific average over all repetitions at the temporal midpoint. The advantage hereof is that it generates a data point for each utterance individually instead of each context pair and thus provides a context-independent measure of coarticulation. Coarticulation was measured at two positions in /ə/, at /ə/ midpoint and at /ə/ boundary, defined as onset (anticipatory) or offset (carryover) of /ə/, which were acoustically identified as the first or last glottal pulse. The two yielded the same pattern of results, although a direct comparison of the two versions of the measure in terms of sensitivity was not possible due to the small sample size ($N = 7$ female adult speakers; Kim et al., 2018).

Articulatory Positioning: Tongue Shape Ratio

Instead of a distance measure based on tongue contours or flesh points, Zharkova, Gibbon, and Hardcastle (2015) quantified coarticulation as the vowel context ratios of five different measures of tongue shape (curvature degree,

curvature position, Dorsum Excursion Index, Tongue Constraint Position Index, and LOC_{a-i} , a tongue bunch location index, which is further explained below; see also Ménard, Aubin, Thibeault, & Richard, 2012; Zharkova, 2013; see Table 16). The main purpose of their study was to compare ultrasound data collection with and without head stabilization (i.e., the ultrasound scanner mounted on a headset or handheld). The results indicate that tongue shape measure LOC_{a-i} is the most robust, as it was the only measure that was not affected by the absence of stabilization. LOC_{a-i} captures the extent of tongue front and tongue back excursion and is calculated as the ratio of tongue height at 1/3 and 2/3 of the length of the tongue curve (measured from the tip). Higher values correspond to a more /i/-like tongue shape, and lower values correspond to a more /a/-like tongue shape (Zharkova et al., 2015).

The LOC_{a-i} tongue shape ratio measure can be seen as the articulatory equivalent of acoustic F2/second moment ratios (see the F2 ratios and First Moment Ratios section) and are suitable for consonant–vowel (CV) or vowel-to-vowel (əCV) anticipatory coarticulation, albeit specifically designed for /i/ and /a/ vowel contexts. Task and elicitation procedures are similar to the mean distance across set/context measure (Zharkova et al., 2011, 2012; see Table 16).

With respect to the comparison between head-mounted or handheld ultrasound recording, the results from Zharkova et al. (2015) indicated that it was possible to collect reliable data without head mount in adolescents ($N = 10$; 13-year-olds). As the authors note, however, this might not hold for younger children. Until it has been conclusively proven to be reliable, it is advised to collect data with head stabilization when investigating coarticulation in younger children.

Articulatory Positioning: Coarticulation Degree

Another measure of coarticulation that has been used in recent ultrasound studies with children is coarticulation

Table 16. Methodological details: tongue shape ratio (LOC_{a-i}; Zharkova et al., 2015).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | Six repetitions of CV syllables consisting of a consonant (/p/, /t/, /s/, and /f/) followed by a vowel context (/i/ and /a/) |
| (2) Tasks used to elicit those targets | Reading/picture naming (text + image on screen) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in carrier phrase (“It’s a...Pam”) |
| (4) The measures obtained from those responses | i/a ratio on the LOC _{a-i} measure of tongue shape |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | No |

Note. CV = consonant–vowel.

degree (Noiray et al., 2018; Rubertus & Noiray, 2018), which can be seen as the articulatory variant of the locus equations metric (see the Locus Equation Metric section; see Table 17). Similarly, coarticulation degree captures whether the positioning of an articulator during the production of a speech sound varies systematically depending on its position in the vowel context by means of a regression analysis. Unlike the acoustics-based equivalent, however, the articulatory measure was used not only for consonant–vowel (CV) anticipatory (Noiray et al., 2018) but also for vowel-to-vowel (VC_v) carryover coarticulation (Rubertus & Noiray, 2018). Articulatory positioning was based on the highest point of the tongue body (horizontally) at the (acoustically determined) temporal midpoint of the segments of interest. Specifically, they measured whether tongue body height in the consonant and /ə/ varied systematically depending on the vowel by regressing the horizontal position of the highest point of the tongue body at C and V midpoint and /ə/ and V midpoint, respectively (see Table 17). Differences in coarticulation degree were expressed in regression coefficients, where a larger value (i.e., a steeper slope) indicates more coarticulation.

Inappropriate Prosody, Especially in the Realization of Lexical or Phrasal Stress

Background

Prosody; Lexical and Phrasal Stress

Prosody is difficult to define and may encompass different aspects of speech for different researchers and clinicians. For present purposes, we will not discuss the many different views of prosody but instead attempt to delineate the aspects of prosody that have received attention in the literature on AOS. To help delineate this domain, we will follow Shriberg and Kent (2013) in using the term *prosody* to refer to suprasegmental aspects of the speech signal that affect the linguistic or communicative structure of an utterance, such as stress, intonation, and pauses (see also Gerken & McGregor, 1998). Excluded from this definition and discussion are paralinguistic, suprasegmental aspects of speech that primarily provide information about the speaker or the speaking context, such as voice quality and overall loudness.

It is important to recognize that there is significant cross-linguistic variation in prosodic structure and the

Table 17. Methodological details: coarticulation degree (Noiray et al., 2018; Rubertus & Noiray, 2018).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | Six repetitions of C1VC2/ə/ pseudowords, V consisting of the tense long vowels /i:/, /y:/, /e:/, /u:/, and /o:/ and C consisting of /b/, /d/, /g/, and /z/ with C1V a fully crossed set and C2 different from C1 |
| (2) Tasks used to elicit those targets | Imitation of prerecorded model |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items preceded by an article (“eine...”) |
| (4) The measures obtained from those responses | Coarticulation degree: mean within stimulus distance in tongue body position in V and C1 midpoint (Noiray et al., 2018) and V and /ə/ midpoint (Rubertus & Noiray, 2018) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | No |
| (7) Norm or reference data available? | Reference data: Graphic displays of regression slope estimates are available for vowel-to-/ə/ carryover coarticulation in consonant contexts /b/, /d/, and /g/ for 3-year-olds ($n = 19$), 4-year-olds ($n = 14$), 5-year-olds ($n = 14$), 7-year olds ($n = 15$), and adults ($n = 13$; $M_{\text{age}} 23$; seven females and six males) with typical development (Rubertus & Noiray, 2018) |

developmental trajectory to acquire adultlike control of prosodic aspects of speech (e.g., Gerken & McGregor, 1998; Kehoe, 2001; Kehoe, Stoel-Gammon, & Buder, 1995). We focus here primarily on English, as most published research on CAS has involved English-speaking children, but the measures reviewed here are expected to be applicable, with appropriate modifications, to other languages as well. In addition, as the focus of the literature with respect to prosody in CAS has been primarily on lexical and phrasal stress, we restrict our discussion to these aspects here as well.

Stress in the linguistic sense refers to perceptual prominence of a syllable in a sequence of syllables and, as such, is a relative phenomenon (it makes little sense to talk of “stress” for monosyllabic utterances). Perceptual prominence of a syllable may involve manipulation of three main perceptual parameters, namely, length (physical attribute: duration), loudness (physical attribute: intensity), and pitch (physical attribute: fundamental frequency or F0). Syllables that are longer, louder, and higher in pitch (or involve a greater pitch change) are perceived as stressed compared to syllables that are shorter, less loud, and lower in pitch (or involve a smaller pitch change). Although these parameters can each signal stress more or less independently (e.g., Patel, 2003; Patel & Campellone, 2009), speakers typically manipulate these aspects in some coordinated fashion. Furthermore, articulatory aspects, such as vowel quality, can also affect the perception of stress (Velleman & Shriberg, 1999). In physical terms, production of prosody involves manipulation of the respiratory system, the phonatory system, and the supralaryngeal (articulatory) system and requires complex coordination across these systems (e.g., Goffman & Malin, 1999). As such, given the changes in vocal tract anatomy during childhood (e.g., Vorperian et al., 2009), it is to be expected that these complex coordination demands can pose difficulties for children with speech motor planning and/or programming impairments.

Lexical stress refers to the stress patterns of individual lexical items, regardless of their sentential context (e.g., *potato* has stress on the second syllable, *pyramid* has stress on the first syllable). Phrasal stress refers to stress patterns in larger, multiword utterances and may serve to highlight important information such as content words and new information (relative to function words or given [old] information) or to indicate a contrast with previous statements or information (e.g., *KRAmer ate the soup* [not Elaine]).

Syllables are grouped into higher level groupings called *metrical feet*, which constitute the domain of stress representation and which themselves are organized into superordinate structures, such as prosodic words (e.g., Kehoe, 2001). Metrical feet contain one or two syllables, and the two most common basic foot types consisting of two syllables are the trochee and the iamb. Trochaic feet have stress on the first syllable (e.g., *mother*, *baby*, *wobble*), and iambic feet have stress on the second syllable (e.g., *balloon*, *hotel*, *forget*). Stressed syllables are sometimes indicated with “S” (strong), and unstressed syllables are indicated with “w” (e.g., *wobble* = Sw and *balloon* = wS). The basic foot pattern in English and other Germanic languages

such as Dutch and German is trochaic (Sw). Most two-syllable words are trochees. Sequences of more than two syllables typically consist of single-syllable feet and trochaic feet. Unfooted unstressed syllables (syllables that do not form part of a trochaic foot) are more vulnerable to omission than footed syllables, both in the course of development (e.g., *banana* [wSw] is more likely to be reduced to *nana* [Sw] than to *bana* [wS]) and in colloquial adult speech (e.g., *opossum* [wSw] is often reduced to *possum* [Sw], not *oposs* [wS]). A number of more specific explanations have been put forward to explain for the patterns of syllable omission in children’s speech, but discussion of these proposals is beyond the scope of this tutorial (see, e.g., Gerken, 1996; Kehoe, 2001, for further discussion).

Prosody and Stress in Typical Development

Few normative data are available for prosodic development, but there appears to be general agreement that syllable omissions should be rare or infrequent by age of 3 or 4 years (Gerken & McGregor, 1998; Kehoe, 2001), and stress errors are considered rare in typical development (Kehoe, 2001). It is important to keep in mind, however, that, in linguistically oriented accounts of prosody and its development, much research has relied on perceptual (transcription-based) methods (e.g., Gerken, 1996; Kehoe, 2000). Given well-known limitations of perceptually based measures (e.g., Goffman, Heisler, & Chakraborty, 2006; Maas & Mailend, 2012), normative suggestions based on such measures must be viewed with some caution. More sensitive measures such as acoustic or kinematic measures may reveal greater insight into prosodic abilities of children. For instance, using acoustic temporal measures (Carter & Gerken, 2004) showed that 2-year-old children who omit syllables (perceptually) do mark the underlying presence of the omitted syllable by lengthening the preceding syllable. Conversely, Goffman (1999, 2004), using kinematic measures, showed that even 4- to 6-year-old children differ from adults in their differentiation between different rhythmic patterns.

Prosody and Stress in Children With CAS

Since the initial descriptions of AOS by Darley and his colleagues (e.g., Darley, Aronson, & Brown, 1975), abnormal prosody has remained a prominent and common feature associated with AOS in children and adults, in scientific investigations and in clinical practice (Ballard, Robin, McCabe, & McDonald, 2010; Caruso & Strand, 1999; Duffy, 2005; Forrest, 2003; Hall, 2000; McCabe, Rosenthal, & McLeod, 1998; Odell, McNeil, Rosenbek, & Hunter, 1991; Odell & Shriberg, 2001; Rosenbek & Wertz, 1972; Shriberg, Aram, & Kwiatkowski, 1997b, 1997c; Strand, McCauley, Weigand, Stoeckel, & Baas, 2013; Velleman & Shriberg, 1999; Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006; Yoss & Darley, 1974). The CAS Technical Report lists the “realization of lexical or phrasal stress” (ASHA, 2007, p. 4) as the core observation of atypical prosody in CAS. More specifically, children with CAS produce less differentiation between

stressed and unstressed syllables (e.g., Munson, Bjorum, & Windsor, 2003; Shriberg et al., 1997b, 1997c, 2003), providing the listener with the impression of equalized stress across syllables or misplaced stress. Changing the prosodic structure of a word by omitting syllables has also been reported (e.g., Velleman & Shriberg, 1999). These differences in realization of lexical or phrasal stress are more readily observed in utterances with iambic feet (Munson et al., 2003; Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003)—the less common grouping of stressed and unstressed syllables in Germanic languages.

Less accurate stress production in children with CAS compared to children without communication impairments and children with other types of communication impairments (e.g., speech delay and phonological impairments) is a consistent finding in the literature, although the observed differences have not always reached statistical significance (e.g., Munson et al., 2003) nor do all children with CAS show abnormal stress production (e.g., Shriberg et al., 1997c). Furthermore, in studies where many different speech and language variables were collected and analyzed as potential indicators for CAS diagnosis, perceived errors in stress marking were the most accurate predictors of expert judgments of CAS diagnosis (prediction accuracy of up to 80%; Murray et al., 2015) and the most successful basis for discriminating children with suspected CAS from children with speech delay (Shriberg et al., 1997b). Nevertheless, the fact that not all children with CAS show abnormal stress production and group differences do not always reach statistical significance suggests that there may be subtypes of CAS (cf. Shriberg et al., 1997c) and/or that perceptual measures of stress production may not be sufficiently robust or sensitive. It is also relevant to note that some studies report relatively poor correspondence between perceptual and acoustic measures of stress production (e.g., Munson et al., 2003).

Perceptual Measures

Percentage Correct Stress

Several studies have examined the adequacy of lexical stress as perceived by listeners to establish a diagnostic basis for identifying children with CAS; see Table 18. Since stress (i.e., a stressed syllable) is a relative construct only identified in the context of nonstressed syllables, the stimuli to elicit a spoken response must include minimally two syllables. These syllables, in turn, may be embedded into longer utterances, such as multisyllabic words, carrier phrases, or sentences. It is important to include both trochaic and iambic feet in the stimuli, because atypical stress production may be more evident in iambic (non)words (e.g., Munson et al., 2003), and including only trochees may therefore not provide an opportunity to observe differences. At the same time, if children also differ in the production of trochaic stress pattern, it may be an indication of a more severe impairment.

The choice of stimuli partially dictates the task that can be used to elicit the responses, with more numerous options for word stimuli. While both words and

pseudowords can be elicited via repetition (Munson et al., 2003; Skinder, Strand, & Mignerey, 1999) or reading (for older children; e.g., Ballard et al., 2010; van Rees, Ballard, McCabe, Macdonald-D'Silva, & Arciuli, 2012), pictures (Murray et al., 2015) and toys can be used to prompt the production of words in a naming task or in a play context or conversation (Odell & Shriberg, 2001¹; Skinder, Connaghan, Strand, & Betz, 2000). Hence, words can be used to collect information about how children store and access linguistic information about stress, whereas for pseudowords, the stress pattern has to be provided in the stimuli (whether spoken or written). Pseudowords do, however, allow for a more straightforward control over different psycholinguistic variables and the phonetic makeup of the words. As stress production may interact with articulatory difficulties (Munson et al., 2003), these aspects may be important to control in order to isolate the effect of stress from the effect of other variables.

Typically, a listener or a group of listeners provide a binary judgment—correct or incorrect marking of stress—against the glossary/recording of the target words. For example, Murray et al. (2015) elicited polysyllabic words from children with a suspected CAS diagnosis in a picture-naming task (the Single-Word Test of Polysyllables; Gozzard, Baker, & McCabe, 2006; see Table 18). To measure the adequacy of stress production, listeners assessed whether the observed stress pattern matched the expected one for the particular word, which was then converted into a percentage of stress matches. Other studies have included additional codes to provide more detail about the nature of the stress production errors. For example, Skinder et al. (2000) asked listeners to judge children's productions of bisyllabic words, including both iambic and trochaic utterances. The listeners identified the productions as (a) misplaced, (b) correct, or (c) equal in terms of stress.

Acoustic Measures

Background

Stress is expressed mainly by three perceptual parameters (length, pitch, and loudness). The physical correlates of these parameters (duration, fundamental frequency or F0, and amplitude/intensity) can be measured acoustically. Several studies have taken advantage of the acoustic approach to study and quantify production of prosody, particularly lexical stress in children with CAS. Acoustic

¹The Prosody-Voice Screening Profile developed by Shriberg, Kwiatkowski, and Rasmussen (1990) is based on multidimensional judgments of the appropriateness of stress based on 24 eligible utterances derived from conversational speech. This system is not further discussed here given the considerable data reduction and perceptual training required (Odell & Shriberg, 2001): "Codors learn to discriminate each prosody-voice (PV) code by training practice that includes learning the perceptual criteria for each code and listening to several hundred audio-taped exemplars obtained from samples of child and adult speakers representing a wide spectrum of speech disorders, including speakers with motor speech disorders" (p. 287).

Table 18. Methodological details: Single-Word Test of Polysyllables (Gozzard et al., 2006).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | 50 Multisyllabic real words: three ($n = 37$), four ($n = 12$), five ($n = 1$) syllables |
| (2) Tasks used to elicit those targets | Picture naming |
| (3) Conditions in which responses are elicited | Quiet, no time pressure |
| (4) The measures obtained from those responses | Percent stress matches (based on binary judgments of match/mismatch) |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Percent stress matches among the two most discriminative measures in distinguishing children with CAS from children without CAS (77%–80% accuracy; Murray et al., 2015) <i>Reliability:</i> Interrater reliability = 91.2% (Murray et al., 2015) |
| (7) Norm or reference data available? | No reference data from children with typical speech. Reference data of children without CAS but with other speech disorders ($N = 15$), $M = 67.3\%$, $SD = 22.4\%$, and children with CAS ($N = 28$), $M = 9.8\%$, $SD = 9.1\%$ (Murray et al., 2015) |

Note. CAS = childhood apraxia of speech.

studies use similar tasks and stimuli as perceptual studies, but some considerations are particularly pertinent in the context of the acoustic approach. We will first consider the effect of phonetic context in stressed and unstressed syllables. Listeners in a perceptual study have the advantage of using all the different cues of stress simultaneously. This allows listeners to weigh different cues differently, depending on the phonetic context. For example, in English, vowels are produced with a longer duration in a syllable with a voiced coda compared to a voiceless one. The acoustic measures of duration, amplitude, and F0 are obtained in isolation. The duration difference that results from different phonetic context or from stress will interact with one another and cannot be separated. In order to interpret the duration change as an effect of stress production, it is important that the phonetic context be controlled for a reliable comparison of stressed and unstressed syllables. In addition, the acoustic measures typically depend on a reliable identification of the nucleus of a syllable (the vowel or a syllabic consonant). The nuclei that are surrounded by stop consonants can be more reliably identified compared to those surrounded by liquids or glides.

In addition to phonetic context, acoustic analysis is particularly sensitive to variables related to phrase- or sentence-level prosodic factors such as phrase-final lengthening and citation intonation. Like phonetic context, these variables affect the same acoustic variables as stress and therefore interact with stress. For example, the last stressed syllable in the final foot of a phrase is subject to phrase-final lengthening. This lengthening can mask the effect of stress on duration in case of trochees and inflate the effect in case of iambs. Using a carrier phrase (e.g., “It’s a [stimulus] again.”) may help to circumvent this issue. Carrier phrases will also help to avoid citation intonation that people often use in picture-naming or single-word reading tasks (Ballard, Djaja, Arciuli, James, & van Doorn, 2012), where people raise their F0 in the end of the utterance as if requesting feedback. Shriberg et al. (2003) also reported that children were playfully varying the duration of the last syllable in

iambs and spondees, which led the authors to exclude these items from analysis.

The acoustic measures are typically obtained from the nucleus of a syllable, and they include the duration of the segment, peak intensity, and peak F0. Sometimes, measures that relate to the timing of the peak F0 and/or amplitude are also included. The magnitude of stress is reflected in the comparisons of these measures between stressed and unstressed syllables—greater difference reflects more pronounced production of stress. Several different techniques have been developed to compare the acoustic measures of stressed and unstressed syllables.

Some studies have compared the raw values of duration, amplitude, and F0 between stressed and unstressed syllables (Nijland, Maassen, Van der Meulen, Gabreëls, et al., 2003; Skinder et al., 2000). For example, Nijland, Maassen, Van der Meulen, Gabreëls, et al. (2003) found a difference in the duration of unstressed syllables in iambic feet when children with CAS were compared to TD children. More specifically, while the duration of stressed syllables was comparable between the two groups, the authors found that children with CAS did not have shorter durations for unstressed syllables in these utterances. The shortcoming of comparing raw values of acoustic measures, such as syllable duration, is that this approach does not take into account individual variation in these measures between different groups. Children with CAS, for example, may have a decreased speaking rate compared to TD children. This systematic difference may interact with the duration differences that are related to stress.

Lexical Stress Ratio

Shriberg et al. (2003) proposed the lexical stress ratio (LSR), an approach to quantify stress production that takes advantage of acoustic correlates of stress see Table 19. The LSR combines duration, intensity, and F0 into one composite score of stress. More specifically, the LSR is the sum of the ratios of three acoustic measures (frequency area under pitch contour trace, amplitude area under

Table 19. Methodological details: lexical stress ratio (LSR; Shriberg et al., 2003).

| Materials and methods | |
|---|---|
| (1) Stimuli or targets being analyzed | Eight bisyllabic real-word trochees (eight iambs and eight spondees excluded from analysis; see Shriberg et al., 2003) |
| (2) Tasks used to elicit those targets | Imitation (from recorded audio model) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure; items in isolation |
| (4) The measures obtained from those responses | Weighted average of ratios of frequency area, amplitude area, and vowel duration |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> No clear pattern of correspondence between LSR values and clinical perceptual judgments of abnormal stress <i>Reliability:</i> Interjudge differences: amplitude = 0.9–1.3 dB; F0 = 10.9–14.0 Hz; duration = 16–18 ms |
| (7) Norm or reference data available? | No reference data of children with typical speech. Range of LSR values for children with (non-CAS) speech delay: 0.65–1.14 (Shriberg et al., 2003) |

Note. CAS = childhood apraxia of speech.

rectified waveform contour trace, and duration) weighted by a constant. Although Hosom, Shriberg, and Green (2004) automated the calculation of LSR using automatic speech recognition, this indicator has not been widely used. While LSR assigns different weight for various acoustic domains combining them into one indicator of stress much like a human listener, it lacks a similar flexibility. Human listeners may weigh different perceptual cues of stress differently, depending on the phonetic and intonational context, while the weights are constant in LSR.

Pairwise Variability Index

Finally, another measure that has been used to quantify stress production in children with and without CAS (Ballard et al., 2012, 2010; Shriberg, Jakielski, & El-Shanti, 2008) is the pairwise variability index (PVI; Low, Grabe, & Nolan, 2000; see Table 20). This index is calculated for each acoustic measure related to stress assignment (duration, intensity, and F0) separately, and it normalizes for the individual variability of speakers for these measures. PVI is calculated by the following formula (from Ballard et al., 2010): $PVI_{(dur)} = ((d_k - d_{k+1}) / ((d_k - d_{k-1}) / 2)) \times 100$, where d is the duration of the k th syllable (see Table 20). This formula illustrates the calculation of PVI for duration; the same formula can be used to calculate the PVI for other acoustic measures by replacing the duration with the measure of interest (e.g., intensity or F0).

Findings to date using the PVI indicate that this measure can reveal differences between speakers with and without AOS (in children and adults). For example, Shriberg et al. (2008) used the PVI to investigate timing and stress characteristics in the speech of three siblings with CAS using the PVI and found a significantly poorer score in one of the three affected speakers, compared to their age-matched controls. With respect to adults with AOS, Vergis et al. (2014) analyzed lexical stress contrastiveness in polysyllabic words produced in isolation and in a carrier sentence, produced by individuals with AOS + aphasia (AOS; $n = 9$), aphasia only ($n = 8$), and unaffected speakers ($n = 8$). The

PVI was used to measure normalized relative vowel duration and peak intensity over the first two syllables of the polysyllabic words. The results showed that speakers with AOS had lower PVI vowel duration values for words with weak–strong stress produced in the sentence condition, compared to controls and individuals with aphasia, and was primarily attributed to disproportionately long vowels in the word-initial weak syllable for AOS participants. Similar findings were reported by Courson et al. (2012). Together, these findings demonstrate that the PVI might be a promising acoustic diagnostic tool in assessing dysprosody in AOS. Ballard et al. (2010) have further demonstrated that the PVI is strongly correlated with perceptual ratings of prosody.

The PVI has several advantages over other approaches of quantifying stress production in addition to normalizing for individual differences of the measures of interest. First, a study by Ballard et al. (2012) provides reference data for the PVI of duration, amplitude, and F0 in a cohort of 73 TD 3- to 7-year-old children. The authors used a picture-naming task to elicit polysyllabic words with Sw and wS stress patterns. According to the results, stress production was adultlike for words with a Sw stress pattern (e.g., “butterfly”) already by age of 3 years. In contrast, even the older children in this cohort differed from adults in their stress production of words with a wS stress pattern (e.g., “potato”), at least with respect to duration and amplitude. These results are crucial for interpreting the stress production differences in children with CAS as they suggest that not all differences in stress production of wS words are a reason for concern in a 5-year-old while differences in words with Sw pattern may be reflective of a delay or disorder.

Second, Ballard et al. (2010) argue that analyzing different correlates of stress separately from one another is a strength of the PVI approach because it may provide a clinician with indications as to which aspect of stress production is most impaired in children with CAS and/or which aspect of stress is the best target in therapy. For example, Ballard et al. examined the effects of therapy, which emphasized only durational contrasts in stress production. While all

Table 20. Methodological details: pairwise variability index (PVI; e.g., Ballard et al., 2010, 2012).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | Multisyllabic nonwords (Ballard et al., 2010) or words (Ballard et al., 2012) |
| (2) Tasks used to elicit those targets | Reading (Ballard et al., 2010) or picture naming (Ballard et al., 2012) |
| (3) Conditions in which responses are elicited | Quiet, no time pressure, in isolation |
| (4) The measures obtained from those responses | PVI for duration, amplitude, and/or F0 |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Strong correlations with perceptual ratings of prosody for three children with CAS (Ballard et al., 2010) <i>Reliability:</i> Intraclass correlation coefficients: .905 (duration), .996 (F0; Ballard et al., 2012); interjudge Pearson $r = .98$ (duration); average difference, 1.22 ms (Ballard et al., 2010) |
| (7) Norm or reference data available? | Reference data available for ages 3–7 years and adults (Ballard et al., 2012). No cutoff values specified, but means and SD available |

Note. CAS = childhood apraxia of speech.

participants with CAS ($n = 3$) showed improvement on the duration contrast, the contrast between other variables, such as intensity and F0, also improved.

With respect to the validity and reliability of the PVI, the following should be noted. At present, validation of the PVI as a measure of dysprosody in CAS is limited to the strong correlations between PVI and perceptual ratings of prosody in three children with CAS, as reported by Ballard et al. (2010). Clearly, further validation using (much) larger samples is needed, in particular, also to validate the PVI as a potential diagnostic marker for CAS (e.g., validation against other measures such as standardized maximum performance tasks). Also, in order to obtain PVI values, it is necessary to divide the stimuli in vocalic and intervocalic intervals based on acoustic information available from the waveform and spectrogram. As noted by White, Liss, and Dellwo (2011), variations exist in the approach of researchers to determine vocalic and consonantal information, although it is not clear whether or to what

extent these measurement differences have an effect on the final result (Liss et al., 2009). In terms of reliability, Ballard and colleagues (Ballard et al., 2012, 2010) reported high Pearson and intraclass correlation coefficients and small interrater differences for their (small) sample, suggesting that PVI measures can be reliably obtained (see Table 20).

Articulatory Measures

Kinematic Pairwise Variability Index

Articulatory measures of prosody may include kinematic measures of movement amplitude and movement duration (e.g., Goffman, 1999, 2004; Grigos & Patel, 2007, 2010; Kopera & Grigos, 2019; see Table 21). Similar to acoustic measures, ratios of duration or amplitude and PVI can be computed to express the degree of differentiation between stressed and unstressed syllables. All the same caveats and considerations as discussed previously, with respect to kinematic measures, apply here as well. To date,

Table 21. Methodological details: kinematic pairwise variability index (PVI; Kopera & Grigos, 2019).

| Materials and methods | |
|---|--|
| (1) Stimuli or targets being analyzed | Two-syllable sequence <i>puppy</i> extracted from multisyllabic nonword <i>puppypop</i> |
| (2) Tasks used to elicit those targets | Cloze sentence or response to question |
| (3) Conditions in which responses are elicited | In story context with prop, no time pressure, in isolation |
| (4) The measures obtained from those responses | PVI for jaw movement duration and amplitude |
| Scientific basis | |
| (5) Standardized measurement protocol? | No |
| (6) Validity and reliability of outcome measures? | <i>Validity:</i> Not reported, except to the extent that the PVI movement duration distinguished children with CAS from TD children, and children were designated as having CAS based on perceived presence of prosodic abnormalities. <i>Reliability:</i> Not reported |
| (7) Norm or reference data available? | Only group means and standard deviations: PVI movement duration: TD = 33.1 (29.3), SSD = 21.9 (18.3), CAS = 18.0 (30.4) PVI movement amplitude: TD = 86.9 (62.0), SSD = 66.3 (45.5), CAS = 73.4 (65.8) |

Note. CAS = childhood apraxia of speech; TD = typically developing; SSD = speech sound disorder.

only one study has applied kinematic measures to study prosody in CAS (Kopera & Grigos, 2019). Kopera and Grigos examined PVI based on movement duration and PVI based on movement amplitude (in addition to acoustic measures) in seven children with CAS, eight children with other SSDs, and nine TD children. Children produced the target word *puppypop* in cloze sentence or in response to a prompt (see Table 21). Kopera and Grigos observed that PVI based on movement duration from perceptually accurate *puppy(pop)* utterances distinguished children with CAS from TD children, and that other children with SSDs did not differ either from children with CAS or TD children. PVI based on movement amplitude did not differ between any groups. Interestingly, PVIs based on acoustic measures (duration, F0 peak, F0 average) did not differ between groups, suggesting that kinematic measures of PVI may be more sensitive to group differences.

Discussion and Conclusions

This tutorial gives an overview of measurement techniques for the assessment of childhood speech motor disorders, in particular CAS, organized according to three levels of directness of measurement—perceptual, acoustic, and kinematic—as well as three symptom domains that are generally considered core deficits of CAS—inconsistency, lengthened and disrupted coarticulation, and inappropriate prosody. Below, the merits of these measures for diagnosis and assessment of underlying deficits are discussed from a broader perspective. In addition, future directions for research and clinical practice are discussed.

The three levels of directness refer to the extent to which the measure directly reflects speech movements. This should not be interpreted to mean that more movement-oriented measures also more directly reflect the underlying deficit. First, this highly depends on the domain of assessment, since some lend themselves better for kinematic measurement (e.g., inconsistency) than others (e.g., inappropriate prosody). Second, for a full description of the clinical aspects of a speech disorder, the three levels are complementary. Especially Kent (2004) but also others (e.g., Brumbach & Goffman, 2014; Goffman, 2010; Kleinow & Smith, 2006; Kloth, Janssen, Kraaimaat, & Brutten, 1995; A. Smith, Goffman, Sasisekaran, & Weber-Fox, 2012) have stressed the role of the interface between language and speech processes in determining typical and deficient speech, arguing that speech movements or speech gestures are perceptually controlled, goal-oriented, and driven by cognition (mental representations). That is, by moving the articulators, the speaker intends to produce an acoustic signal that can be understood by a listener; in addition to the reaction of the listener, the speaker's own monitoring of the acoustic result thereby forms a crucial criterion to determine success. From many studies, it has become clear that there is no one-to-one relationship between speech movements and perceptual result. According to Guenther, Hampson, and Johnson (1998), the invariant targets of speech are not vocal tract shapes, but regions in auditory perceptual space,

which implies that the acoustic–auditory results serve as primary reference frame for the control and monitoring of speech movements; as a consequence, there can be large variability in movement given a particular auditory goal. Thus, to fully understand the control of speech movements and to reveal the underlying deficits in speech disorders, the three levels of directness are complementary.

From the perspective of clinical evaluation and intervention, some characteristics are better suited to be described at the perceptual level, especially phonemic errors and prosody; others at the acoustic level, especially phonetic distortions, coarticulation, and prosody; and still others at the kinematic level, especially coarticulation, stability, and gestural coordination. The first reason why this is the case resides in the limitations of our knowledge of the relation between auditory, acoustic, and articulatory phonetics. If for a particular sound we do not precisely know all the acoustic and kinematic correlates of a correct production, the perceptual judgment of the clinician outperforms acoustic and kinematic measures in determining correctness. However, the second, clinically more important reason to adopt a multilevel approach is the notion of reference frames presented in the previous paragraph. Both for the clinician and for the speaker, an auditory approach to correct deviant articulatory movements can be the primary diagnostic and therapeutic channel. For instance, the optimal instruction for correcting vowel productions in a person with dysarthria who has difficulty with tongue elevation might very well be a perceptual one: “produce the vowel more /i/-like.” To what extent a sound is /i/-like may be easier to assess and easier to control at the perceptual than at the kinematic level: “elevate your tongue body closer to the palatal–alveolar region.” The acoustic level can be of help to focus on particular aspects of the auditory percept and to link the auditory percept to kinematic aspects. To illustrate these relations, we give an example. Technical developments between 1970 and 1990 allowed for constructing equipment that visualizes acoustic characteristics of speech. Thus, the so-called *visual speech apparatus* was designed for the deaf and hard of hearing, providing a visual feedback of the produced vowel in acoustic F1–F2 space (Povel & Arends, 1991). Here, the instruction could be “try to move the vowel-dot into the upper-right corner of this vowel triangle.” Similar acoustics-based feedback systems have been developed since, to give enhanced feedback for speech disorders (e.g., McAllister Byun & Hitchcock, 2012; McAllister Byun, Swartz, Halpin, Szeredi, & Maas, 2016; Shuster, Ruscello, & Toth, 1995) or second-language acquisition (e.g., Bliss, Abel, & Gick, 2018; Dowd, Smith, & Wolfe, 1998; Hirata, 2004; Lambacher, 1999).

In addition to the different levels of reference frames, a reason why the levels of directness need to be analyzed in a complementary fashion is that speech is produced in a larger context—words, utterances, dialogues—in which higher levels of control interact with lower levels. A process-oriented approach requires analyses of not only the end product, that is, speech movements, but also the processes of conceptualization, formulation, encoding, planning, and

control involved in producing those movements. These processes run off in a cascade-like fashion, such that all processes are simultaneously active, working on different parts of the utterance: the higher the level, the more advanced the preparation. The simultaneous organization may require a trade-off in attention allocation. Fluent speakers may focus their attention on the message they want to get across and to a lesser extent on formulating eloquent sentences but leave all articulation processes to the automatic pilot. For people with language and speech disorders, speaking may be more like a dual or even triple or quadruple task, in which case much more attention must be allocated to the formulation and articulation processes as well. Thus, in this view, processing levels interact, and speech symptoms must therefore be studied in context.

The same two principles (context dependency and interaction between levels) apply to each of the three CAS characteristics: inconsistency, lengthened and disrupted coarticulation, and inappropriate prosody. These can be approached with each of the three levels of measurement, and none of the levels is better or more direct than any of the others across contexts. However, the approach chosen determines to a large extent the data that are collected and thus the interpretation that can be given regarding the underlying deficit. Comprehensive studies are needed that include more than one diagnostic feature and more than one level of measurement.

For whom was this tutorial written? First of all, for researchers who can spend time to consider alternative methods and have the resources to implement those that are judged to be optimal. There is no holy grail; different research questions and different study populations require different methods. The past decades were a time of technical development that allowed for powerful acoustic analyses and saw the emergence of fine-grained kinematic measurement procedures. The line of research and technical development has been focused on finding acoustic and kinematic correlates of perceptual phenomena, as well as the other way around: finding perceptual effects of kinematic and acoustic phenomena. Now, to move beyond objectifying perceptual phenomena with acoustic and kinematic measures, we need to adopt an integrated approach in which all levels are included and all levels are interpreted in what they have to offer for diagnosis and treatment.

Second, for clinicians, who generally need to work with what they have. Being better informed about diverse methods of assessment fosters an analytic view on underlying processes and alternative methods of listening, watching, and measuring. There are quite a few assessment protocols around; most of these are not validated and used primarily based on pragmatic considerations with respect to time and equipment available. Most of the cited studies do not have the scope and the volume, in terms of numbers of participants or scope of measurements, to yield norm data or even reference data that could be applied outside the context of the particular study and thus could lead to generalizations. This implies that no measure so far has proven to have clear

diagnostic value on its own. A first step to improve this situation is to adopt a more analytic, process-oriented way of theorizing about measurements and their relation with underlying deficits in speech disorders. The second step is to conduct clinical research to come up with validated consensus measurement protocols to operationalize, quantify, and eventually standardize assessments, so that we can better compare children across studies—and interpret observations from individual clients in a clinical setting—using replicable methods.

References

- Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75–86.
- American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* (Report No. TR2007-00278). Rockville, MD: Author. Retrieved from <http://www.asha.org/policy>
- Anderson, A., Lowit, A., & Howell, P. (2008). Temporal and spatial variability in speakers with Parkinson's disease and Friedreich's ataxia. *Journal of Medical Speech-Language Pathology*, 16(4), 173.
- Aram, D. M., & Horwitz, S. J. (1983). Sequential and non-speech praxic abilities in developmental verbal apraxia. *Developmental Medicine & Child Neurology*, 25, 197–206.
- Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics*, 14(2), 131–150.
- Ballard, K. J., Djaja, D., Arciuli, J., James, D. G., & van Doorn, J. (2012). Developmental trajectory for production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research*, 55(6), 1822–1835.
- Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53(5), 1227–1245.
- Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M., & Perkell, J. (2013, September). *Speech planning as an index of speech motor control maturity*. Paper presented at the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon, France.
- Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4), 591–627.
- Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, 69(1), 231–238.
- Betz, S. K., & Stoel-Gammon, C. (2005). Measuring articulatory error consistency in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 19(1), 53–66.
- Bickley, C. (1986). Formant estimation of high fundamental frequency speech. *The Journal of the Acoustical Society of America*, 79(S1), S38.
- Bliss, H., Abel, J., & Gick, B. (1998). Computer-assisted visual articulation feedback in L2 pronunciation instruction. *Journal of Second Language Pronunciation*, 4(1), 129–153.
- Bradford, A., & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics & Phonetics*, 10(2), 77–101.

- Brancazio, L., & Fowler, C. A.** (1998). On the relevance of locus equations for production and perception of stop consonants. *Perception and Psychophysics*, 60(1), 24–50.
- Brumbach, A. C. D., & Goffman, L.** (2014). Interaction of language processing and motor skill in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 57(1), 158–171.
- Burt, L., Holm, A., & Dodd, B.** (1999). Phonological awareness skills of 4-year-old British children: An assessment and developmental data. *International Journal of Language & Communication Disorders*, 34(3), 311–335.
- Butcher, A.** (1989). Measuring coarticulation and variability in tongue contact patterns. *Clinical Linguistics & Phonetics*, 3(1), 39–47.
- Butcher, A., & Weiher, E.** (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, 4(1), 59–74.
- Carter, A., & Gerken, L.** (2004). Do children's omissions leave traces? *Journal of Child Language*, 31(3), 561–586.
- Caruso, A. J., & Strand, E. A.** (1999). *Clinical management of motor speech disorders in children*. New York, NY: Thieme.
- Chang, S.-E., Ohde, R. N., & Conture, E. G.** (2002). Coarticulation and formant transition rate in young children who stutter. *Journal of Speech, Language, and Hearing Research*, 45(4), 676–688.
- Chau, T., Young, S., & Redekop, S.** (2005). Managing variability in the summary and comparison of gait data. *Journal of Neuro-Engineering and Rehabilitation*, 2(1), 22.
- Chitoran, I., Goldstein, L., & Byrd, D.** (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 419–447). Berlin, Germany: Mouton de Gruyter.
- Cho, T.** (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 32(2), 141–176.
- Cleland, J., McCron, C., & Scobbie, J. M.** (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics & Phonetics*, 27(4), 299–311.
- Courson, M.-E. A., Ballard, K. J., Canault, M., Layfield, C. A., Scholl, D. I., & Gentil, C.** (2012). Lexical stress production in healthy and apraxic speakers of Australian English or French. *Journal of Medical Speech-Language Pathology*, 20(4), 47–53.
- Cummins, F., Lowit, A., & van Brenk, F.** (2014). Quantitative assessment of interutterance stability: Application to dysarthria. *Journal of Speech, Language, and Hearing Research*, 57(1), 81–89.
- Daniloff, R., & Hammarberg, R.** (1973). On defining coarticulation. *Journal of Phonetics*, 1(3), 239–248.
- Darley, F. L., Aronson, A. E., & Brown, J. R.** (1975). *Motor speech disorders*. Philadelphia, PA: Saunders.
- De Jong, K. J.** (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, 97(1), 491–504.
- Dodd, B.** (1995). Procedures for classification of subgroups of speech disorder. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 49–64). San Diego, CA: Singular.
- Dodd, B.** (2005). *Differential diagnosis and treatment of children with speech disorder* (2nd ed.). London, United Kingdom: Whurr.
- Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A.** (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London, United Kingdom: The Psychological Corporation.
- Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A.** (2009). *Diagnostic Evaluation of Articulation and Phonology—US Edition (DEAP)*. San Antonio, TX: Pearson.
- Dowd, A., Smith, J., & Wolfe, J.** (1998). Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech*, 41(1), 1–20.
- Duffy, J. R.** (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Mosby.
- Edwards, J., Beckman, M. E., & Fletcher, J.** (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, 89(1), 369–382.
- Ekelman, B. L., & Aram, D. M.** (1984). Spoken syntax in children with developmental verbal apraxia. In D. M. Aram (Ed.), *Seminars in speech and language: Assessment and treatment of developmental apraxia* (pp. 97–110). New York, NY: Thieme-Stratton.
- Farnetani, E., & Recasens, D.** (1997). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 371–404).
- Feng, Y., & Max, L.** (2014). Accuracy and precision of a custom camera-based system for 2-D and 3-D motion tracking during speech and nonspeech motor tasks. *Journal of Speech, Language, and Hearing Research*, 57(2), 426–438.
- Forrest, K.** (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology*, 12(3), 376–380.
- Forrest, K., Dinnsen, D. A., & Elbert, M.** (1997). Impact of substitution patterns on phonological learning by misarticulating children. *Clinical Linguistics & Phonetics*, 11(1), 63–76.
- Forrest, K., Elbert, M., & Dinnsen, D. A.** (2000). The effect of substitution patterns on phonological treatment outcomes. *Clinical Linguistics & Phonetics*, 14(7), 519–531.
- Forrest, K., & Iuzzini-Seigel, J.** (2008). A comparison of oral motor and production training for children with speech sound disorders. *Seminars in Speech and Language*, 29(4), 304–311.
- Fowler, C. A.** (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics*, 55(6), 597–610.
- Gerken, L.** (1996). Prosodic structure in young children's language production. *Language*, 72, 683–712.
- Gerken, L., & McGregor, K.** (1998). An overview of prosody and its role in normal and disordered child language. *American Journal of Speech-Language Pathology*, 7(2), 38–48.
- Gibbon, F., Hardcastle, W., & Nicolaidis, K.** (1993). Temporal and spatial aspects of lingual coarticulation in /k/ sequences: A cross-linguistic investigation. *Language and Speech*, 36(2–3), 261–277.
- Gibbon, F., & Lee, A.** (2007). Electropalatography as a research and clinical tool. *Perspectives on Speech Science and Orofacial Disorders*, 17(1), 7–13.
- Gibson, T., & Ohde, R. N.** (2007). F2 locus equations: Phonetic descriptors of coarticulation in 17- to 22-month-old children. *Journal of Speech, Language, and Hearing Research*, 50(1), 97–108.
- Goffman, L.** (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research*, 42(6), 1499–1517.
- Goffman, L.** (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research*, 47(5), 1088–1102.

- Goffman, L.** (2010). Dynamic interaction of motor and language factors in normal and disordered development. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 137–152). Oxford, United Kingdom: Oxford University Press.
- Goffman, L., Heisler, L., & Chakraborty, R.** (2006). Mapping of prosodic structure onto words and phrases in children's and adults' speech production. *Language and Cognitive Processes*, 21(1–3), 25–47.
- Goffman, L., & Malin, C.** (1999). Metrical effects on speech movements in children and adults. *Journal of Speech, Language, and Hearing Research*, 42(4), 1003–1015.
- Gozzard, H., Baker, E., & McCabe, P.** (2006). Children's productions of polysyllables. *ACQuiring Knowledge in Speech, Language and Hearing*, 8(3), 113–116.
- Graetzer, S., Fletcher, J., & Hajek, J.** (2015). Locus equations and coarticulation in three Australian languages. *The Journal of the Acoustical Society of America*, 137(2), 806–821.
- Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W.** (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43(1), 239–255.
- Green, J. R., Nip, I. S. B., & Maassen, B.** (2010). Some organization principles in early speech development. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 171–188). Oxford, United Kingdom: Oxford University Press.
- Grigos, M. I.** (2009). Changes in articulator movement variability during phonemic development: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, 52(1), 164–177.
- Grigos, M. I., & Kolenda, N.** (2010). The relationship between articulatory control and improved phonemic accuracy in childhood apraxia of speech: A longitudinal case study. *Clinical Linguistics & Phonetics*, 24, 17–40.
- Grigos, M. I., Moss, A., & Lu, Y.** (2015). Oral articulatory control in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 58(4), 1103–1118.
- Grigos, M. I., & Patel, R.** (2007). Articulator movement associated with the development of prosodic control in children. *Journal of Speech, Language, and Hearing Research*, 50(1), 119–130.
- Grigos, M. I., & Patel, R.** (2010). Acquisition of articulatory control for sentential focus in children. *Journal of Phonetics*, 38(4), 706–715.
- Guenther, F. H., Hampson, M., & Johnson, D.** (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105(4), 611–633.
- Guyette, T. W., & Diedrich, W. M.** (1981). A critical review of developmental apraxia of speech. In N. J. Lass (Ed.), *Speech and language. Advances in basic research and practice* (pp. 1–49). New York, NY: Academic Press.
- Hall, P. K.** (2000). A letter to the parent(s) of a child with developmental apraxia of speech. Part I: Speech characteristics of the disorder. *Language, Speech, and Hearing Services in Schools*, 31, 169–172.
- Hardcastle, W. J.** (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4(1–3), 247–263.
- Hardcastle, W. J., & Tjaden, K.** (2008). 32 Coarticulation and speech impairment. *The handbook of clinical linguistics*, 506–524.
- Harrington, J.** (2010). Acoustic phonetics. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *A handbook of phonetics* (Vol. 116, pp. 81–129). Oxford, United Kingdom: Wiley-Blackwell.
- Heisler, L., Goffman, L., & Younger, B.** (2010). Lexical and articulatory interactions in children's language production. *Developmental Science*, 13(5), 722–730.
- Hertrich, I., & Ackermann, H.** (1995). Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech*, 38(2), 159–187.
- Hertrich, I., & Ackermann, H.** (1999). Temporal and spectral aspects of coarticulation in ataxic dysarthria: An acoustic analysis. *Journal of Speech, Language, and Hearing Research*, 42(2), 367–381.
- Hirata, Y.** (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17(3–4), 357–376.
- Holm, A., Crosbie, S., & Dodd, B.** (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language & Communication Disorders*, 42(4), 467–486.
- Hosom, J.-P., Shriberg, L., & Green, J. R.** (2004). Diagnostic assessment of childhood apraxia of speech using automatic speech recognition (ASR) methods. *Journal of Medical Speech-Language Pathology*, 12(4), 167.
- Howell, P., Anderson, A. J., Bartrip, J., & Bailey, E.** (2009). Comparison of acoustic and kinematic approaches to measuring utterance-level speech variability. *Journal of Speech, Language, and Hearing Research*, 52(4), 1088–1096.
- Howell, P., Anderson, A. J., & Lucero, J.** (2010). Motor timing and fluency. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 215–225). Oxford, United Kingdom: Oxford University Press.
- Ingram, D.** (2002). The measurement of whole-word productions. *Journal of Child Language*, 29(4), 713–733.
- Iuzzini-Seigel, J.** (2012). *Inconsistency of speech in children with childhood apraxia of speech, phonological disorders, and typical speech development* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Iuzzini-Seigel, J., & Forrest, K.** (2010). Evaluation of a combined treatment approach for childhood apraxia of speech. *Clinical Linguistics & Phonetics*, 24(4–5), 335–345.
- Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R.** (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research*, 60(5), 1194–1210.
- 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.
- Iuzzini-Seigel, J., Hogan, T. P., Rong, P., & Green, J. R.** (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development*, 3(1), 53–68.
- Katz, W. F., & Bharadwaj, S.** (2001). Coarticulation in fricative-vowel syllables produced by children and adults: A preliminary report. *Clinical Linguistics & Phonetics*, 15(1–2), 139–143.
- Katz, W. F., Machetanz, J., Orth, U., & Schönle, P.** (1990). A kinematic analysis of anticipatory coarticulation in the speech of anterior aphasic subjects using electromagnetic articulography. *Brain and Language*, 38(4), 555–575.
- Kehoe, M. M.** (2000). Truncation without shape constraints: The latter stages of prosodic acquisition. *Language Acquisition*, 8(1), 23–67.
- Kehoe, M. M.** (2001). Prosodic patterns in children's multisyllabic word productions. *Language, Speech, and Hearing Services in Schools*, 32(4), 284–294.
- Kehoe, M. M., Stoel-Gammon, C., & Buder, E. H.** (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38(2), 338–350.

- Kent, R. D.** (2004). Models of speech motor control: Implications from recent developments in neurophysiological and neuro-behavioral science. In B. Maassen, R. Kent, H. F. M. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 1–28). Oxford, United Kingdom: Oxford University Press.
- Kent, R. D., & Minifie, F. D.** (1977). Coarticulation in recent speech production models. *Journal of Phonetics*, 5(2), 115–133.
- Kewley-Port, D.** (1982). Measurement of formant transitions in naturally produced stop consonant–vowel syllables. *The Journal of the Acoustical Society of America*, 72(2), 379–389.
- Kim, Y., Coalson, G., & Berry, J.** (2018). A kinematic analysis of coarticulation effects on schwa. *Folia Phoniatrica et Logopaedica*, 70(3–4), 203–212.
- Kleinow, J., & Smith, A.** (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 43(2), 548–559.
- Kleinow, J., & Smith, A.** (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology*, 48(4), 275–287.
- Kloth, S., Janssen, P., Kraaimaat, F., & Brutten, G.** (1995). Speech-motor and linguistic skills of young stutterers prior to onset. *Journal of Fluency Disorders*, 20(2), 157–170.
- Kopera, H. C., & Grigos, M. I.** (2019). Lexical stress in childhood apraxia of speech: Acoustic and kinematic findings. *International Journal of Speech-Language Pathology*. <https://doi.org/10.1080/17549507.2019.1568571>
- Krull, D.** (1988). *Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops*. Sweden: Department of Linguistics, University of Stockholm.
- Krull, D.** (1989). Second formant locus patterns and consonant–vowel coarticulation in spontaneous speech. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm*, 10, 87–108.
- Kühnert, B., Hoole, P., & Mooshammer, C.** (2006). *Gestural overlap and C-center in selected French consonant clusters*. Paper presented at the 7th International Seminar on Speech Production (ISSP), Ubatuba, Brazil.
- Kühnert, B., & Nolan, F.** (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 1–30). Cambridge, United Kingdom: Cambridge University Press.
- Lambacher, S.** (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12(2), 137–156.
- Lee, S., Potamianos, A., & Narayanan, S.** (1999). Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105, 1455.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M.** (1967). Perception of the speech code. *Psychological Review*, 74(6), 431.
- Lindblom, B.** (1963). On vowel reduction (Report No. 29). Stockholm, Sweden: The Royal Institute of Technology, Speech Transmission Laboratory.
- Lisker, L., & Abramson, A. S.** (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.
- Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N.** (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, 52(5), 1334–1352.
- Low, E. L., Grabe, E., & Nolan, F.** (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377–401.
- Lucero, J. C.** (2005). Comparison of measures of variability of speech movement trajectories using synthetic records. *Journal of Speech, Language, and Hearing Research*, 48(2), 336–344.
- Lucero, J. C., Munhall, K. G., Gracco, V. L., & Ramsay, J. O.** (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*, 40(5), 1111–1117.
- Lundeborg, I., Nordin, E., Zeipel-Stjerna, M., & McAllister, A.** (2015). Voice onset time in Swedish children with phonological impairment. *Logopedics Phoniatrics Vocology*, 40(4), 149–155.
- Maas, E., & Farinella, K. A.** (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 55(2), 561–578.
- Maas, E., & Mailend, M.-L.** (2012). Speech planning happens before speech execution: Online reaction time methods in the study of apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 55(5), 1523–1534.
- Maas, E., & Mailend, M.-L.** (2017). Fricative contrast and coarticulation in children with and without speech sound disorders. *American Journal of Speech-Language Pathology*, 26(2S), 649–663.
- Maassen, B., Nijland, L., & Terband, H.** (2010). Developmental models of childhood apraxia of speech. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 243–258). Oxford, United Kingdom: Oxford University Press.
- Maassen, B., Nijland, L., & Van der Meulen, S.** (2001). Coarticulation within and between syllables by children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 15(1–2), 145–150.
- Macrae, T., Tyler, A. A., & Lewis, K. E.** (2014). Lexical and phonological variability in preschool children with speech sound disorder. *American Journal of Speech-Language Pathology*, 23(1), 27–35.
- Marion, M. J., Sussman, H. M., & Marquardt, T. P.** (1993). The perception and production of rhyme in normal and developmentally apraxic children. *Journal of Communication Disorders*, 26(3), 129–160.
- Marquardt, T. P., Jacks, A., & Davis, B.** (2004). Token-to-token variability in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, 18(2), 127–144.
- Marquardt, T. P., Sussman, H. M., Snow, T., & Jacks, A.** (2002). The integrity of the syllable in developmental apraxia of speech. *Journal of Communication Disorders*, 35, 31–49.
- McAllister Byun, T., & Hitchcock, E. R.** (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*, 21, 207–221.
- McAllister Byun, T., Swartz, M. T., Halpin, P. F., Szeredi, D., & Maas, E.** (2016). Direction of attentional focus in biofeedback treatment for /r/ misarticulation. *International Journal of Language & Communication Disorders*, 51, 384–401.
- McCabe, P., Rosenthal, J. B., & McLeod, S.** (1998). Features of developmental dyspraxia in the general speech-impaired population? *Clinical Linguistics & Phonetics*, 12(2), 105–126.
- McIntosh, B., & Dodd, B. J.** (2008). Two-year-olds’ phonological acquisition: Normative data. *International Journal of Speech-Language Pathology*, 10(6), 460–469.
- Mefferd, A.** (2015). Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis. *Journal of Speech, Language, and Hearing Research*, 58(3), 576–589.

- Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica*, 64(2), 64–72.
- Ménard, L., Cathiard, M.-A., Dupont, S., & Tiede, M. (2013). Anticipatory lip gestures: A validation of the movement expansion model in congenitally blind speakers. *The Journal of the Acoustical Society of America*, 133(4), EL249–EL255.
- Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E. (2004). An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics*, 32(3), 291–312.
- Moss, A., & Grigos, M. I. (2012). Interarticulatory coordination of the lips and jaw in childhood apraxia of speech. *Journal of Medical Speech-Language Pathology*, 20(4), 127.
- Munson, B., BJORUM, E. M., & WINDSOR, J. (2003). Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language, and Hearing Research*, 46(1), 189–202.
- Murray, E., Iuzzini, J., Maas, E., Terband, H., & Ballard, K. (2018). *Diagnosis of childhood apraxia of speech compared to other speech sound disorders: A systematic review*. Paper presented at the ASHA Convention 2018, Boston, MA, November 15-17, 2018.
- Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 58(1), 43–60.
- Namasivayam, A. K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., . . . van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech. *International Journal of Language & Communication Disorders*, 50(4), 529–546.
- Namasivayam, A. K., Pukonen, M., Goshulak, D., Yu, V. Y., Kadis, D. S., Kroll, R., . . . De Nil, L. F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders*, 46(3), 264–280.
- Nijland, L., Maassen, B., Hulstijn, W., & Peters, H. (2004). Speech motor coordination in Dutch-speaking children with DAS studied with EMMA. *Journal of Multilingual Communication Disorders*, 2(1), 50–60.
- Nijland, L., Maassen, B., Van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16(6), 461–483.
- Nijland, L., Maassen, B., & Van der Meulen, S. (2003). Evidence of motor programming deficits in children diagnosed with DAS. *Journal of Speech, Language, and Hearing Research*, 46(2), 437–450.
- Nijland, L., Maassen, B., Van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 17(1), 1–24.
- Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, 36(5), 959–972.
- Nittrouer, S., Munhall, K., Kelso, J. S., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *The Journal of the Acoustical Society of America*, 84(5), 1653–1661.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32(1), 120–132.
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 61(6), 1355–1368.
- Noiray, A., Cathiard, M.-A., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 153–171). Oxford, United Kingdom: Oxford University Press.
- Noiray, A., Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America*, 129(1), 340–349.
- Noiray, A., Ménard, L., Cathiard, M.-A., Abry, C., & Savariaux, C. (2004, October). *The development of anticipatory labial coarticulation in French: A pioneering study*. Paper presented at the Eighth International Conference on Spoken Language Processing, Jeju Island, South Korea.
- Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America*, 133(1), 444–452.
- Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (in press). Back from the future: Non-linear anticipation in adults and children's speech. *Journal of Speech, Language, and Hearing Research* (this special issue).
- Odell, K. H., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1991). Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. *Journal of Speech and Hearing Research*, 34(1), 67–80.
- Odell, K. H., & Shriberg, L. D. (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics & Phonetics*, 15(4), 275–307.
- Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1), 151–168.
- Patel, R. (2003). Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. *Journal of Speech, Language, and Hearing Research*, 46(6), 1401–1415.
- Patel, R., & Campellone, P. (2009). Acoustic and perceptual cues to contrastive stress in dysarthria. *Journal of Speech, Language, and Hearing Research*, 52(1), 206–222.
- Pollock, K. E., & Hall, P. K. (1991). An analysis of the vowel misarticulations of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 5, 207–224.
- Povel, D.-J., & Arends, N. (1991). The visual speech apparatus: Theoretical and practical aspects. *Speech Communication*, 10(1), 59–80.
- Preston, J. L., & Koenig, L. L. (2011). Phonetic variability in residual speech sound disorders: Exploration of subtypes. *Topics in Language Disorders*, 31(2), 168.
- Preston, J. L., Molfese, P. J., Gumkowski, N., Sorcinelli, A., Harwood, V., Irwin, J. R., & Landi, N. (2014). Neurophysiology of speech differences in childhood apraxia of speech. *Developmental Neuropsychology*, 39(5), 385–403.
- Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis*. New York, NY: Springer.
- Recasens, D. (2004). The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). *Journal of Phonetics*, 32(3), 435–453.

- Recasens, D., & Pallarès, M. D. (2001). Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics*, 29(3), 273–301.
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, 102(1), 544–561.
- Rosenbek, J. C., & Wertz, R. T. (1972). A review of fifty cases of developmental apraxia of speech. *Language, Speech, and Hearing Services in Schools*, 3(1), 23–33.
- Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PLOS ONE*, 13(9), e0203562.
- Sadagopan, N., & Smith, A. (2008). Developmental changes in the effects of utterance length and complexity on speech movement variability. *Journal of Speech, Language, and Hearing Research*, 51(5), 1138.
- Schumacher, J. G., McNeil, M. R., Vetter, D. K., & Yoder, D. E. (1986, November). *Articulatory consistency and variability in apraxic and non-apraxic children*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Detroit, MI.
- Sharf, D. J., & Ohde, R. N. (1981). Physiologic, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 5, pp. 153–247). New York, NY: Academic Press.
- Shriberg, L. D. (1994). Developmental phonological disorders: Moving toward the 21st century—Forwards, backwards, or endlessly sideways? *American Journal of Speech-Language Pathology*, 3, 26–28.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997a). Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research*, 40(2), 273–285.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997b). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, 40(2), 286–312.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997c). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, 40(2), 313–337.
- Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeney, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics*, 17(7), 549–574.
- Shriberg, L. D., Jakielski, K. J., & El-Shanti, H. (2008). Breakpoint localization using array-CGH in three siblings with an unbalanced 4q;16q translocation and childhood apraxia of speech (CAS). *American Journal of Medical Genetics Part A*, 146A(17), 2227–2233.
- Shriberg, L. D., & Kent, R. D. (2013). *Clinical phonetics* (4th ed.). Boston, MA: Pearson.
- Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The prosody-voice screening profile*. Tucson, AZ: Communication Skill Builders.
- Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., . . . Wilson, D. L. (2015). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker. *Journal of Speech, Language, and Hearing Research*, 60(4), S1096–S1117.
- Shuster, L. I., Ruscello, D. M., & Toth, A. R. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology*, 4, 37–44.
- Siren, K. A., & Wilcox, K. A. (1995). Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *Journal of Speech and Hearing Research*, 38(2), 351–359.
- Skinder, A., Connaghan, K., Strand, E., & Betz, S. (2000). Acoustic correlates of perceived lexical stress errors in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, 8(4), 279–284.
- Skinder, A., Strand, E. A., & Mignerey, M. (1999). Perceptual and acoustic analysis of lexical and sentential stress in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, 7, 133–144.
- Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41(1), 18–30.
- Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of Fluency Disorders*, 37(4), 344–358.
- Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research*, 104(3), 493–501.
- Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1), 22–33.
- Smith, B., Marquardt, T., Cannito, M., & Davis, B. (1994). Vowel variability in developmental apraxia of speech. In J. A. Till, K. M. Yorkston, & D. R. Beukelman (Eds.), *Motor speech disorders: Advances in assessment and treatment* (pp. 81–89). Baltimore, MD: Brookes.
- Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, 70(4), 976–984.
- Song, J. Y., Demuth, K., Evans, K., & Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *The Journal of the Acoustical Society of America*, 133(5), 2931–2946.
- Sosa, A. V., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*, 33(1), 31–50.
- Southwood, M., Dagenais, P., Sutphin, S., & Garcia, J. M. (1997). Coarticulation in apraxia of speech: A perceptual, acoustic, and electropalatographic study. *Clinical Linguistics & Phonetics*, 11(3), 179–203.
- Story, B. H., & Bunton, K. (2016). Formant measurement in children's speech based on spectral filtering. *Speech Communication*, 76, 93–111.
- Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research*, 56(2), 505–520.
- Sussman, H. M. (1994). The phonological reality of locus equations across manner class distinctions: Preliminary observations. *Phonetica*, 51(1–3), 119–131.
- Sussman, H. M., Bessell, N., Dalston, E., & Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *The Journal of the Acoustical Society of America*, 101(5), 2826–2838.
- Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *The Journal of the Acoustical Society of America*, 94(3), 1256–1268.
- Sussman, H. M., Hoemeke, K. A., & McCaffrey, H. A. (1992). Locus equations as an index of coarticulation for place of

- articulation distinctions in children. *Journal of Speech and Hearing Research*, 35(4), 769–781.
- Sussman, H. M., Marquardt, T., & Doyle, J. (2000). An acoustic analysis of phonemic integrity and contrastiveness in developmental apraxia of speech. *Journal of Medical Speech-Language Pathology*, 8(4), 301–313.
- Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, 90(3), 1309–1325.
- Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., & Smith, J. (1996). Consonant–vowel interdependencies in babbling and early words: Preliminary examination of a locus equation approach. *Journal of Speech and Hearing Research*, 39(2), 424–433.
- Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics*, 58(6), 936–946.
- Terband, H. (2017, July). *Deviant coarticulation in childhood apraxia of speech (CAS) does not include hyperarticulation*. Paper presented at the 7th International Conference on Speech Motor Control, Groningen, the Netherlands.
- Terband, H., & Maassen, B. (2010). Speech motor development in childhood apraxia of speech (CAS): Generating testable hypotheses by neurocomputational modeling. *Folia Phoniatrica et Logopaedica*, 62, 134–142.
- Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech (CAS). *Journal of Speech, Language, and Hearing Research*, 52(6), 1595–1609.
- Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2014). Auditory–motor interactions in pediatric motor speech disorders: Neurocomputational modeling of disordered development. *Journal of Communication Disorders*, 47, 17–33.
- Terband, H., Maassen, B., & Maas, E. (2016). Toward a model of pediatric speech sound disorders (SSD) for differential diagnosis and therapy planning. In P. van Lieshout, B. Maassen, & H. Terband (Eds.), *Speech motor control in normal and disordered speech: Future developments in theory and methodology* (pp. 81–110). Rockville, MD: ASHA.
- Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica et Logopaedica*. <https://doi.org/10.1159/000499426>
- Terband, H., Maassen, B., van Lieshout, P., & Nijland, L. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders*, 44(1), 59–74.
- Terband, H., van Zaalen, Y., & Maassen, B. (2012). Lateral jaw stability in children with developmental speech disorders. *Journal of Medical Speech-Language Pathology*, 20(4), 112–118.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Timmins, C., Hardcastle, W., McCann, J., Wood, S., & Wishart, J. (2008). Coarticulation in children with Down's syndrome: An electropalatographic analysis. *Proceedings of ISSP 2008—8th International Seminar on Speech Production (INRIA) Strasbourg, France*, 273–276.
- Tyler, A. A., & Lewis, K. E. (2005). Relationships among consistency/variability and other phonological measures over time. *Topics in Language Disorders*, 25(3), 243–253.
- Tyler, A. A., Lewis, K. E., & Welch, C. M. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology*, 12(3), 289–298.
- Tyler, A. A., Williams, M. J., & Lewis, K. E. (2006). Error consistency and the evaluation of treatment outcomes. *Clinical Linguistics & Phonetics*, 20(6), 411–422.
- van Brenk, F., & Lowit, A. (2012). The relationship between acoustic indices of speech motor control variability and other measures of speech performance in dysarthria. *Journal of Medical Speech-Language Pathology*, 20(4), 24–29.
- van Lieshout, P., & Moussa, W. (2000). The assessment of speech motor behavior using electromagnetic articulography. *The Phonetician*, 81, 9–22.
- van Lieshout, P., & Namasivayam, A. K. (2010). Speech motor variability in people who stutter. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 191–214). Oxford, United Kingdom: Oxford University Press.
- van Rees, L. J., Ballard, K. J., McCabe, P., Macdonald-D'Silva, A. G., & Arciuli, J. (2012). Training production of lexical stress in typically developing children using orthographically biased stimuli and principles of motor learning. *American Journal of Speech-Language Pathology*, 21(3), 197–206.
- Velleman, S. L., & Shriberg, L. D. (1999). Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 42(6), 1444–1460.
- Vergis, M. K., Ballard, K. J., Duffy, J. R., McNeil, M. R., Scholl, D., & Layfield, C. (2014). An acoustic measure of lexical stress differentiates aphasia and aphasia plus apraxia of speech after stroke. *Aphasiology*, 28(5), 554–575.
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., . . . Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, 125(3), 1666–1678.
- Walton, J. H., & Pollock, K. E. (1993). Acoustic validation of vowel error patterns in developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 7, 95–111.
- Wambaugh, J. L., Duffy, J. R., McNeil, M., Robin, D. A., & Rogers, M. A. (2006). Treatment guidelines for acquired apraxia of speech: A synthesis and evaluation of the evidence. *Journal of Medical Speech-Language Pathology*, 14(2), xv–xxxiii.
- Weismer, G., Yunusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria. *Journal of Speech, Language, and Hearing Research*, 46(5), 1247–1261.
- Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18, 3–35.
- White, L., Liss, J. M., & Dellwo, V. (2011). Assessment of rhythm. In A. L. R. D. Kent (Ed.), *Assessment of motor speech disorders* (pp. 231–251). San Diego, CA: Plural.
- Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: A developmental study of motor speech skills in humans. *Neuroscience Letters*, 347(1), 29–32.
- Yao-Tresguerres, S., Iuzzini-Seigel, J., & Forrest, K. (2009). *The effect of therapy on children's use of inconsistent substitutes*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.
- Yoss, K. A., & Darley, F. L. (1974). Developmental apraxia of speech in children with defective articulation. *Journal of Speech and Hearing Research*, 17(3), 399–416.
- Yu, V. Y., Kadis, D. S., Oh, A., Goshulak, D., Namasivayam, A., Pukonen, M., . . . Pang, E. W. (2014). Changes in voice onset time and motor speech skills in children following motor speech therapy: Evidence from /pa/ productions. *Clinical Linguistics & Phonetics*, 28(6), 396–412.

-
- Yunusova, Y., Green, J. R., & Mefferd, A.** (2009). Accuracy assessment for AG500, electromagnetic articulograph. *Journal of Speech, Language, and Hearing Research*, *52*(2), 547–555.
- Zharkova, N.** (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*, *50*(1), 76–81.
- Zharkova, N., Gibbon, F. E., & Hardcastle, W. J.** (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics*, *29*(4), 249–265.
- Zharkova, N., Hewlett, N., & Hardcastle, W. J.** (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, *15*(1), 118–140.
- Zharkova, N., Hewlett, N., & Hardcastle, W. J.** (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, *42*(2), 193–208.
- Ziegler, W., & von Cramon, D.** (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language*, *26*(1), 117–130.