Clinical Linguistics & Phonetics, March 2005; 19(2): 67–87

Taylor & Francis healthsciences

Timing errors in two children with suspected childhood apraxia of speech (sCAS) during speech and music-related tasks

BEATE PETER, & CAROL STOEL-GAMMON

Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd Street, University of Washington, Seattle, WA 98195, USA

(Received 28 July 2003; accepted 25 January 2004)

Abstract

Impaired speech prosody has been identified as a critical feature of suspected childhood apraxia of speech (sCAS). Lexical stress productions of children with sCAS have been characterized as 'excessive/equal/misplaced'. This investigation examines two potential explanations of this particular deficit, articulatory difficulty and impaired intrinsic timing. Two children with a diagnosis of sCAS (ages 4 years, 3 months and 9 years, 5 months) and two age-matched controls were observed during three speech and three music tasks. Acoustic analysis revealed that in all tasks, the performance of the controls was more accurate than that of the children with sCAS. Timing structures and accuracy are discussed with respect to diagnostic status, age, speech and music tasks, and timing unit size.

Keywords: *Suspected childhood apraxia of speech (sCAS), prosody, timing errors, speech, music, acoustic analysis*

Introduction

Current definitions of childhood apraxia of speech

Childhood apraxia of speech (CAS) is a primary speech disorder thought to interfere with the motor programming stages of speech production. Since its inception as a proposed clinical entity by Morley, Court and Miller (1954), it has been the subject of controversy. Varying levels of disagreement have ranged from outright denial of the disorder (e.g. 'a label in search of a population'; Guyette and Diedrich, 1981) to differences of opinion regarding the profile of associated primary and secondary characteristics. Those who affirm the validity of CAS as a clinical entity promote a core set of associated features including a limited phonemic inventory; omission errors; vowel errors; inconsistent articulation errors; altered suprasegmental characteristics such as disordered prosody, voice quality, and fluency; increased errors on longer units of speech output; difficulty imitating words and phrases; predominant use of simple syllable shapes; impaired volitional oral movements;

Correspondence: Beate Peter, Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, University of Washington, Seattle, WA 98195, USA. E-mail: bypeter@u.washington.edu

ISSN 0269-9206 print/ISSN 1464-5076 online
DOI: 10.1080/02699200410001669843

reduced expressive vs. receptive language skills; and reduced diadochokinetic rates (Davis, Jakielski and Marquardt, 1998). Aware of the controversy associated with the label 'childhood apraxia of speech', the present authors qualify it with 'suspected', using the acronym sCAS.

With Shriberg, Aram and Kwiatkowski's (1997a, b, c) attempt to isolate a diagnostic marker for sCAS, the exact nature of prosodic errors in this population has moved to a more prominent place in the sCAS discussion. This group of researchers observed a higher incidence of errors related to lexical stress in the speech of children with sCAS, compared to typically developing children. Errors were analysed using the Prosody-Voice Screening Profile (PVSP; Shriberg, Kwiatkowski and Rasmussen, 1990) and coded with an error label called 'Excessive/Equal/Misplaced', a category combining the percepts of 'excessively forceful, punctuated monostress, misplaced word stress and/or sound blocks or prolongations' (McSweeney and Shriberg, 2001) under a single error code. Shriberg *et al.* (1990) suggested that stress production errors could be a diagnostic marker of sCAS or one of its subtypes. In a metrical analysis of sub-samples from that study, Velleman and Shriberg (1999) found that stress error patterns were not qualitatively different from those observed in typically developing children, but that these stress errors occurred to a greater degree and persisted longer. In light of this reanalysis, stress-related errors may serve to distinguish children with sCAS from typically developing children based on quantitative rather than qualitative measures. McSweeney and Shriberg (2001) suggest that this classification based on stress errors serves as a potential diagnostic marker of at least one subtype of sCAS. In an attempt to examine the underlying phonological concept of syllabicity, Marquardt, Sussman, Snow and Jacks (2002) reported that, in their sample of six participants, the three children with sCAS showed a poorer perception and/or representation of syllable structure than their three typically developing peers.

Motivation for examining extralinguistic manifestations of timing deficits

Shriberg and colleagues view the prosodic dimension of sCAS as an inherently speech related phenomenon (Shriberg *et al.*, 1997a, b, c; Velleman and Shriberg, 1999). Using a different theoretical framework, Alcock, Passingham, Watkins and Vargha-Khadem (2000) investigated prosodic errors in an extralinguistic domain, namely in the domain of music, by analysing the pitch and timing behaviours during speech and music-related tasks of family members diagnosed with 'verbal dyspraxia' and comparing it to the performance of age-matched controls. They found that timing abilities in the individuals with the disorder were affected during both speech and music tasks while pitch abilities were intact during both types of activities. This result has two implications: (1) it suggests that prosodic errors in sCAS may have an extralinguistic dimension that manifests in non-speech tasks such as in music, and (2) it validates an approach to investigating prosodic errors in children with sCAS that focuses on timing rather than on other acoustic correlates of stress, such as intensity and intonation. One hypothesis underlying the present study hence suggests that prosodic errors in children with sCAS may be based on impaired timing mechanisms that also manifest in non-speech tasks. This 'Internal Metronome Hypothesis' provides the present study's motivation to investigate extralinguistic manifestations of timing deficits in this population, using acoustic measures.

Theoretical support for the Internal Metronome Hypothesis comes from a variety of behavioural and neuroimaging studies. Regarding behavioural observations, most individuals are able to tap two harmonically related rhythms (e.g. 1:2; 1:3; 1:4) separately with two hands without difficulty, but experience interference with non-harmonic ratios ratios such as 2:3; 3:4; 3:5 (Schmidt and Lee, 1999). Polyrhythmic interferences occur crossmodally, e.g. during timed speech tasks while performing timed hand tapping where the two pulse rates form a non-harmonic ratio (Klapp, 1981). Another example is the observed difficulty of tapping a pulse sequence while reciting a nursery rhyme (Peters, 1977). These behavioural observations are consistent with the notion that one central time keeper underlies rhythm production across limbs and modalities. Regarding neuroimaging data, Wong (2002) summarized a substantial number of such studies investigating speech prosody and argued that, in addition to linguistic levels and task-specific characteristics, durational unit size may be an important factor in determining neural processing systems. Gandour (1987) suggested that larger linguistic units, e.g. words and phrases, are preferentially processed in right hemisphere structures while the left hemisphere specializes in syllabic and sub-syllabic units. Zatorre (2001) made analogous suggestions with respect to neural processing of musical stimuli, implying that right hemisphere structures are assigned fine spectral but broad temporal analysis tasks, with left hemisphere structures performing complementarily.

Besides drawing support from Alcock *et al.*'s (2000) results, a timing-based approach to studying the speech prosody of children with sCAS is justified in light of the observation that vowel durations are a critical feature associated with lexical stress. Ramus, Nespor and Mehler (2000) showed that quantity and variability of vocalic vs. consonantal segments are more efficient predictors of stress-timed, syllable-timed, and mora-timed percepts than traditional constructs of isochrony. Building on this result, Low, Grabe and Nolan (2000) showed that a quantity variable derived from the durational difference of any two adjacent vowel nuclei, the 'Pairwise Variability Index', is an even more sensitive tool than mere vowel durations in differentiating gradations along a continuum ranging from stress timing to mora timing. Since many studies of prosody impairments in children with sCAS focus on lexical stress, duration measures, especially those of vowel nuclei, appear to be highly appropriate. Furthermore, both of these studies imply that perceptual judgements are insufficient to capture temporal fine structures such as syllable component durations, and point to the need for acoustic analyses.

To return to the existing body of literature on prosodic deficits in sCAS, most studies of prosodic deficits in sCAS focus on deficits during speech tasks exclusively. Systematic investigations of errors during music tasks have not been published. Similarly, studies isolating the fine structure of timing in the speech of these children are scarce. Thirdly, most data analyses reported in the literature are based on perceptual judgements. According to Nijland, Maassen, Van der Meulen, Gabreels, Kraaimaat and Schreuder R. (2002), acoustic analyses of the speech output from children with sCAS are rare, and, subsequently, little is known about the phonetic characteristics of this speech. Exceptions to the second and third shortcomings are Skinder, Strand and Mignerey (1999), Skinder, Connaghan, Strand and Betz (2000), Green, Shriberg and Campbell (2002), and Munson, Bjorum and Windsor (2003), all of which validate an instrumental, timing-based approach. In particular, Skinder *et al.* (1999) found that the stress productions of five children with sCAS in bisyllabic and multisyllabic words and in sentences were perceived as less accurate than those of five control participants, and that the children with sCAS used the same acoustic parameters as the control participants did, but with more intersubject variability. Skinder *et al.* (2000) correlated perceptual judgements and acoustic parameters in the bisyllabic word productions of children with sCAS and found differences for acoustic correlates and the perceptual codes for peak fundamental frequency and amplitude only. Green *et al.* (2002) measured variability of speech event and pause event durations in the speech of children with sCAS and found that their speech event durations were less variable than those produced by typical controls and controls with phonologic disorder, an acoustic finding consistent with the percept of ‘staccato’-like speech; the pause events, on the other hand, were more variable than those in the two control groups. Comparing the bisyllabic non-word productions of five children with sCAS against those of five children with phonologic delay (PD), Munson *et al.* (2003) found that the participants with sCAS marked the differences between stressed and unstressed syllables in similar ways as their peers with PD, but that their perceptual accuracy was lower than that of their PD peers.

Research questions

The purposes of the present study were to (1) test the validity and precision of the current sCAS definition by investigating whether children with this diagnosis have identifiable timing difficulties, and (2) test the Internal Metronome Hypothesis by comparing timing features in their speech output to those in their music output. The following questions were addressed:

1. Are the timing structures in the speech of children with sCAS less accurate and less adult-like than those produced by typically developing (TD) children?
2. Does the degree of temporal accuracy in the speech of children with sCAS have a direct correlate in the domain of music productions?

Method

Participants

Two male children with suspected childhood apraxia of speech (sCAS), ages 4;3 and 9;5 (years;months) and two typically developing (TD), female children, ages 4;3 and 8;9, participated in this study. The gender disparity between the participants with sCAS and the TD participants was not intended per study design; rather, it was a result of the type of participants who responded to the recruitment process. The acronyms sCAS-Y, TD-Y, sCAS-O and TD-O are used to refer to the younger child with sCAS, the younger TD child, the older child with sCAS and the older TD child, respectively. The two children with sCAS were referred by local speech-language pathologists and the TD children were recruited through a notice at a local paediatrician’s office. To recruit the children with sCAS, speech-language pathologists were asked to forward the study information to parents of children who exhibited at least eight out of 11 characteristics identified by Davis, Jakielski and Marquardt (1998; listed previously). The presence of eight of these characteristics in the participants referred with a sCAS diagnosis was confirmed by the first author.

All participants met the following inclusionary criteria: (1) monolingual, English-speaking home environment, (2) hearing screening passed at 25 dB SPL at 0.5, 1, 2, 4 and 8 kHz, (3) normal receptive language as measured by the *Test of Auditory Comprehension of Language–Revised* (*TACL–3*; Carrow-Woolfolk, 1999), (4) normal cognitive functioning as estimated by the *Peabody Picture Vocabulary Test, Version B* (*PPVT-B*; Dunn and Dunn, 1997) and (5) normal oral structures and functioning, as measured with the motor control score in a sub-set of tasks from the *Verbal Motor Production Assessment for Children*

(VMPAC; Hayden and Square, 1999; items 14 through 19; items 33 through 37). For the TD children, additional inclusionary criteria were (1) age-appropriate articulation skills, as measured by the *Goldman Fristoe Test of Articulation 2 (GFTA-2*; Goldman and Fristoe, 2000) and by listener judgement of normal intelligibility, and (2) age-appropriate expressive language skills, as measured by either the recalling sentences in context from the *Clinical Evaluation of Language Fundamentals–Preschool* (*CELF–P*; Wiig, Secord and Semel, 1992) or the recalling sentences sub-test from the *Clinical Evaluation of Language Fundamentals–3* (*CELF–3*; Semel, Wiig and Secord, 1995), as appropriate for age. Table I contains a summary of data obtained from testing for inclusionary criteria and from parent comments.

Instrumentation

Data were collected in a quiet clinic room in the Department of Speech and Hearing Sciences at the University of Washington and involved two sessions of approximately

Table I. Participant characteristics related to inclusionary criteria, plus other, parent-supplied information. Participants are listed by diagnostic status and age (years;months)

Type of Information	sCAS (4;3)	TD (4;3)	sCAS (9;5)	TD (8;9)
Standard Score: *Test of Auditory Comprehension of Language 3*	11	12	8	12
Standard Score: *Peabody Picture Vocabulary Test, Version B*	94	107	104	118
Standard Score: Goldman-Fristoe Test of Articulation 2	91	112	<40	103
Standard Score: *Recalling Sentences in Context* sub-test from the *Clinical Evaluation of Language Fundamentals – Preschool* or *Recalling Sentences* from the *Clinical Evaluation of Language Fundamentals – 3*	7	8	10	13
Handedness	Left	Right	Right	Right
History of music instruction	Group 1 hr/week since age 6 mo	Group 45 min/ week, age 6 to 18 mo	Per typical school curriculum	Per typical school curriculum
History of speech therapy	Since age 30 mo Currently 1 hr/week	n/a	Since age 30 mo Currently 2.5 hr/week	n/a
Observed features from Davis *et al.*'s (1998) sCAS profile	Limited repertoire, omissions, inconsistency, altered prosody, more errors with increased complexity, simple syllable shapes, reduced expressive language, reduced diadochokinetic rates	n/a	Limited repertoire, inconsistency, altered prosody, more errors with increased complexity, difficulty imitating words and phrases, simple syllable shapes, reduced expressive language, reduced diadochokinetic rates	n/a

1 hour each. The children's caregivers were reimbursed for their participation plus parking expenses. Video-recordings were obtained using a digital Sony 3CCD camcorder. Sound was recorded with a Sony PCM M1digital audiotape recorder. Participants were seated at a table of comfortable height, with the microphone placed on the table, approximately 40 cm from the participant's mouth. Pre-recorded auditory stimuli were presented via CD, using a Panasonic RX-DT610 CD player system. Using CoolEdit Pro 1.2 software (Syntrillium Software Corporation, 1998), the digital audio recordings were redigitized at a sampling rate of 22050 Hz with 16 bit quantization. Acoustic analyses were performed using the Praat 4.0.45 acoustic analysis software (Boersma and Weenink, 2003). Statistical analyses and graphing were performed using Stata Intercooled 7 (Stata Corporation, 2001) and Microsoft Excel (Microsoft Corporation, 2002) software programs.

Tasks and analysis metrics

Sentence imitation. The recalling sentences in context sub-test from the *Clinical Evaluation of Language Fundamentals–Preschool* test (Wiig, Secord and Semel, 1992), used as an inclusionary tool for the younger two participants to rule out language impairments, served a dual role in that it also provided the stimuli for the sentence imitation task. This sub-test consists of a story with embedded dialogue phrases, to be imitated by the examinee directly following their presentation. In order to create consistent stimulus material for all participants, a University of Washington undergraduate student enrolled in the drama department was recorded reading the story. The student had been instructed to produce the dialogue portions of the text in a manner consistent with spontaneous, rather than read, speech. His spoken rendition of the text was recorded on digital audiotape, then reformatted and copied to a CD, from where it was presented to each study participant. For purposes of direct comparisons, only those phrases produced with the correct word sequence by all participants were selected for acoustic analysis, resulting in a sample of seven short phrases with a total of 32 syllables. Because the older participants imitated more of the sentences with an intact word count, a larger sample, consisting of eight longer sentences and 77 syllables, was selected from their productions and analysed separately.

For the acoustic analysis of the sentence imitation and all other speech material in the present study, the measured unit was the duration of the vowel nucleus from each syllable. Vowel nuclei were measured in accordance with the general segmentation guidelines offered by Peterson and Lehiste (1960). Devoiced vowel portions were included in the vowel duration (Low *et al.*, 2000), provided that the formant pattern from the voiced portion was continuous with it. Durational accuracy was assessed with the child:adult correlation coefficient (*r*).

Non-word imitation. The rhythm sub-test from the *Tennessee-Test of Rhythm and Intonation Patterns* (*T-TRIP*; Koike and Asp, 1981) yielded the stimuli for a non-word imitation task. This sub-test, presented to the participants via CD player, contains 14 repetitive sequences of the syllable 'ma' with varied numbers (two to six) of syllables and stress patterns, for example 'ma.MA' or 'ma.ma.MA.ma' where capitalized letters indicate a stressed syllable. Each item provides two opportunities for imitation. For acoustic analysis, each participants' second production was chosen unless the first production had a more accurate syllable count. Items 13 ('MA.<PAUSE>MA.ma') and 14 ('MA.ma.<PAUSE>MA') were excluded from analysis because three of the participants began their imitations at the onset of the embedded pause rather than waiting until the entire syllable string had been presented. Six of the 12 presented sequences contained two stressed syllables. In order to allow for direct comparison, syllables omitted by one or more of the participants were excluded from the analysis. Furthermore, all phrase-final syllables were excluded because equivalent vowel end points for the two types of recording environments (commercially manufactured test stimuli vs. live recording) could not be established reliably. The resulting sample consisted of 26 syllables.

For qualitative information, the types of omitted syllables in the non-word imitations were catalogued. Accuracy was assessed for three aspects of these imitations: number of syllables, vowel nucleus durations, and duration ranges for stressed and unstressed syllables. For the number of produced syllables, a relative error score (number of missing/extraneous syllables divided by the target number) was calculated. For vowel nucleus durations, the child:adult correlation score was calculated. Duration ranges for stressed vs. unstressed target syllables, following the stressed/unstressed coding in the test manual, were evaluated for their mutual distinctiveness, where a more adult-like profile was characterized by two completely separate duration ranges and low variability within each duration range, while a less adult-like profile showed a range overlap and larger spread. The degree of separation and the magnitude of the spread were captured within one variable, effect size, which was calculated using Formula 1 to arrive at a difference between means as a function of the pooled standard deviation:

$$ES = \frac{\bar{X}_{str} - \bar{X}_{unstr}}{(SD_{str} + SD_{unstr})/2} \quad (1)$$

where:

$\bar{X}_{str}$ = mean duration of stressed vowels

$\bar{X}_{unstr}$ = mean duration of unstressed vowels

SD_{str} = standard deviation for stressed vowels

SD_{unstr} = standard deviation for unstressed vowels

Monosyllabic word generation. Participants were presented with a randomized set of picture cards and asked to name each, using the carrier phrase, 'I can say ___'. The pictures represented common nouns and verbs with a variety of onset and rime characteristics. Appendix A contains a list of the pictured items. Three durational variables, onset, nucleus, and coda, were measured and averaged separately for each participant. This task was added to investigate durational characteristics of sub-syllabic components as a function of diagnostic status and age. Accuracy and variability measures were not applied.

Singing Happy Birthday. Each participant was asked to sing the song *Happy Birthday*, following affirmation from the child or the caregiver that the child was familiar with that song. This task was not preceded by a model. The authors acknowledge that singing is a speech task overlaid on a music task and thus cannot be considered to be strictly a music task. It was nevertheless included in the study for the following reasons: (1) pilot testing showed that younger children had substantial difficulty humming the tune and clapping the rhythm in isolation; (2) the duration units in the song are much longer than those in speech and can thus be presumed to drive production; and (3) the older child with sCAS omitted initial consonants freely while singing, even though he produced many of them in speech, (presumably due to therapy), which may be interpreted as indicating that he was thinking more about singing than about speaking.

The unit of measurement for the singing task was the interval between vowel onsets, rather than the interval between syllable onsets, because the older child with sCAS tended to omit onset consonants while singing. For this song, each syllable represents one note (as opposed to stretching one syllable across two or more notes, such as the syllable ‘sleigh’ in the song *Jingle Bells*); thus vowel onset intervals were equivalent to note durations. Because durations in music productions are subject to common, nonlinear time warping effects (Gabrielsson, 1999), mathematical duration relationships derived from musical notation are inadequate as a reference for durational accuracy. Hence, a musically trained adult male, a former school choir director and current University of Washington graduate student, was recorded singing *Happy Birthday* three times. All three of his productions were highly consistent with each other in terms of note durations (all three Pearson’s $r > 0.99$); consequently, one rendition was chosen as the basis of comparison for each child participant’s production. The analysis was limited to comparisons of the first seven measures of the song because of the natural *ritardando* present in the eighth and following measures of many actualizations of the song. In addition to note durations, ratios for any adjacent note durations were calculated.

Accuracy scores in the participants’ productions were derived from child:adult correlations of note durations and $note_x/note_{x+1}$ ratios. Both types of correlations are independent of singing rate, which is important in this unmodelled task. The first of these accuracy measures is global, while the second takes local effects into account. In addition, rhythmic stability throughout the measured portion of the song was estimated using the relative standard deviation (RSD; i.e. the standard deviation/mean ratio) from the eight quarter notes contained in it.

Clapped rhythm imitation. The first author presented each participant with eight short rhythms consisting of two to nine hand claps each. Appendix B shows schematics of the clap sequences. The participant’s task was to imitate each rhythm directly following its presentation. Of the 45 claps and resulting 30 onset-to-onset intervals contained in the combined rhythms modelled for the participants, a subset of ten intervals, consisting of those produced correctly by all four participants, yielded the analysable data. Since the experimenter presented the clap series individually to each child, the consistency of these four models was assessed with pairwise correlations, and correlation coefficients were 0.99, 0.97, 0.98, 0.99, 0.99 and 0.97. Because the younger participants’ productions limited this sample size, the productions of the two older participants were analysed separately, with an intact interval count of 37. The experimenter’s consistency in presenting this expanded clap series was ascertained with a correlation coefficient of 0.99.

Measured units were the intervals between burst onsets, as defined by the earliest visible deviation from silence associated with the burst in the waveform in a window of approximately 8 msec. Accuracy estimates were derived from (1) the overall number of omitted and/or added claps, by calculating a relative error score (number of errored claps/ number of target claps), and (2) clap interval durations, by calculating the child:adult correlation coefficient.

Paced repetitive tapping. Following detailed instructions and a practice trial, the participants listened to a metronome, set at 104 beats/minute, for approximately 20 beats. During the subsequent 30 beats, participants tapped the table, using their preferred hand, attempting to synchronize with the metronome. At the end of this phase, the metronome was turned off and the participants continued tapping at a rate as close to the metronome's as possible, for a minimum of 31 more beats. The task was repeated with rates of 132 and 160 beats/minute.

As with the clapped rhythm imitation task, the measured units for this task were the intervals between burst onsets. Analyses were based on 27 of the child's unaccompanied tap intervals, beginning with the fourth tap after the metronome was turned off. The first three unaccompanied taps were excluded in order to minimize any startle effects. Accuracy for overall rate for each of the three metronome settings was based on the child's mean interval and expressed as a relative error (RE) score, where the absolute difference between the child's mean interval and the target interval was divided by the target interval. This measure indicates how closely the child matched the metronome's interval in general, with a lower RE score reflecting a closer match. The measure for consistency was based on relative standard deviations (RSD, i.e. the ratio of the standard deviation over the mean), where a lower RSD score indicated lower variability and thus, higher accuracy.

Additional measure: mean unit sizes. In order to examine potential associations between accuracy and average durations of the measured units, a mean unit size was calculated for each measured variable, pooling all participants' productions. Accuracy measurements are ordered by task-specific unit size, and trends are reported.

Descriptive Statistics

Because of the small number of participants, traditional hypothesis testing for group differences is inappropriate for this study. Accuracy of performance, as estimated by child:adult correlation coefficients and/or mean relative errors, is profiled for each participant across tasks. By task, each participant's accuracy measure is then compared to that from (1) the same-age peer with the opposite diagnostic status (TD vs. sCAS) and (2) the peer with the same diagnostic category but the opposite age designator (older vs. younger). These comparisons yield the data for a qualitative description of an accuracy advantage by diagnostic status and age. Since multiple datapoints were collected from each of the four participants and the independence assumption is hence not satisfied, probability values for the obtained outcomes cannot be reported.

Reliability

Approximately 19% of the data from all participants and four tasks (non-word imitation, singing *Happy Birthday*, clapped rhythm imitation, paced repetitive tapping) and were measured independently by an undergraduate student in the Department of Speech and Hearing Sciences at the University of Washington. An equivalent proportion of the remaining two tasks was measured by a second student, a graduate student from the same department. Neither student was informed of the participants' diagnostic categories and ages. For each experimental task, the duration measurements were compared to those from the first author, and a proportional difference was computed by dividing the average absolute difference score between the two measured durations by the first author's average measured durations. Deviances of less than 20% (a proportion of 0.2) were considered acceptable for the purposes of this study. Table II lists these average proportions by study task.

Results

Mean unit sizes

For all speech and music tasks, an estimate of the timing unit size in question was obtained from the means and standard deviations of the participants' pooled productions, or, in the case of the paced repetitive tapping task, from the metronome settings. Table III summarizes these unit sizes.

Results from study tasks

Sentence imitation. In this task, all participants imitated stimulus phrases from the *Clinical Evaluation of Language Fundamentals–Preschool* test (Wiig *et al.*, 1992), and a shared sub-set from their productions, consisting of 32 syllables, was analysed. The child:adult correlation

Table II. Data reliability, expressed as the proportion of average measurement differences, relative to the first author's measured duration

Task	Proportional Measurement Difference of Averages
Sentence imitation	0.176
Non-word imitation	0.089
Monosyllabic word generation	0.170
Singing *Happy Birthday*	0.036
Clapped rhythm imitation	0.002
Paced repetitive tapping	0.025

Table III. Summary of timing unit sizes for each study task. Means and standard deviations from participants' pooled productions, plus metronome settings

Task	Measured Unit	Mean (SD) in Seconds
Sentence imitation	Vowel nuclei	0.181 (0.11)
Non-word imitation	Vowel nuclei	Unstressed syllables: 0.128 (0.06) Stressed syllables: 0.253 (0.07)
Monosyllabic word generation	Onsets	0.187 (0.108)
	Vowel nuclei	0.294 (0.149)
	Codas	0.258 (0.160)
Singing *Happy Birthday*	Vowel onset intervals (=notes)	0.567 (0.295)
Clapped rhythm imitation	Clap onset intervals	Smaller dataset: 0.388 (0.139) Larger dataset: 0.465 (0.205)
Paced repetitive tapping	Tap onset intervals	Metronome settings: 0.577 0.455 0.375

(r) indicated that the TD participants were more accurate than their peers with sCAS, and within diagnostic category, the older participants were more accurate than their younger peers. The correlation ranked the participants TD-O (r=0.81), TD-Y (r=0.61), sCAS-O (r=0.59), sCAS-Y (r=0.56), from highest to lowest child:adult association. When the expanded dataset (77 syllables) from the two older participants exclusively was considered, correlation coefficients were 0.85 for TD-O and 0.74 for sCAS-O.

Non-word imitation. This task consisted of imitating sequences of the syllable 'ma' with varied length and stress characteristics. Regarding omission and insertion errors, it is noteworthy that sCAS-Y omitted unstressed syllables preceding a trochee (xx.MA.ma.ma, where xx denotes the omitted element) on two occasions and one unstressed syllable following a trochee (MA.ma.xx). His TD peer omitted no syllables. sCAS-O omitted an unstressed syllable following a trochee (MA.ma.xx) on two occasions, while his TD peer occluded an unstressed syllable following a stressed one once by lengthening the consonant, which is consistent with an incomplete separation of the lips during the vowel (target: MA.ma.ma; production: MA.mma). Overall, the relative error scores computed from the omitted syllables as a ratio of the 45 target syllables were as follows: sCAS-Y: 0.09; TD-Y: 0.00; sCAS-O: 0.04; TD-O: 0.00.

For the acoustic analysis, all phrase-final syllables and those not produced by all participants were excluded, resulting in a subset of 26 syllables. The child:adult correlation coefficients (r) were as follows: TD-O: 0.95; TD-Y: 0.80; sCAS-O: 0.77; and sCAS-Y: 0.38.

When vowel durations were grouped into duration ranges for stressed vs. unstressed syllables in each participant's productions, a complete separation of stressed vs. unstressed duration ranges was apparent for the adult and TD-O. Minimal overlap was observed for TD-Y, while the overlap was substantial for sCAS-O and sCAS-Y. As a measure of distinctiveness between stressed and unstressed vowel durations, effect size was calculated as follows: Adult: 9.8; TD-O: 4.9; TD-Y: 3.6; sCAS-O: 2.3; sCAS-Y: 0.7. Effect size thus indicated a substantially greater distinction between durational ranges for the adult and the two TD participants than for the two participants with sCAS. An additional observation is that the two participants with sCAS produced longer vowel durations in general than their TD peers and the adult, both in terms of means (adult: 0.161 s; TD-O: 0.154 s; sCAS-O: 0.217 s; TD-Y: 0.179 s; sCAS-Y: 0.242 s) and maximum durations (adult: 0.258 s; TD-O: 0.273 s; sCAS-O: 0.411 s; TD-Y: 0.354 s; sCAS-Y: 0.456 s). Figure 1 illustrates the varying degrees of durational distinctiveness, using horizontal dotplots by participant and syllable type.

Monosyllabic word generation. The participants named 50 pictures representing common, monosyllabic words, using a carrier phrase. A subset of 35 words was selected for acoustic analysis, excluding all productions that could not be measured reliably (e.g. words with glide onsets or liquid codas) and all words not produced in an intact fashion by all participants. While this sample contained no phoneme deletions or insertions, substitutions and distortions were evident in the productions of the two participants with sCAS.

Regarding syllable durations as a whole, within each matched pair, the TD child showed shorter durations than the peer with sCAS, and within diagnostic category, the older child showed shorter durations than the younger child. The mean syllable durations were as follows: sCAS-Y: 0.748 s; TD-Y: 0.617 s; sCAS-O: 0.692 s; and TD-O: 0.560 s. This pattern is consistent with increased speaking rates for TD status and greater age.

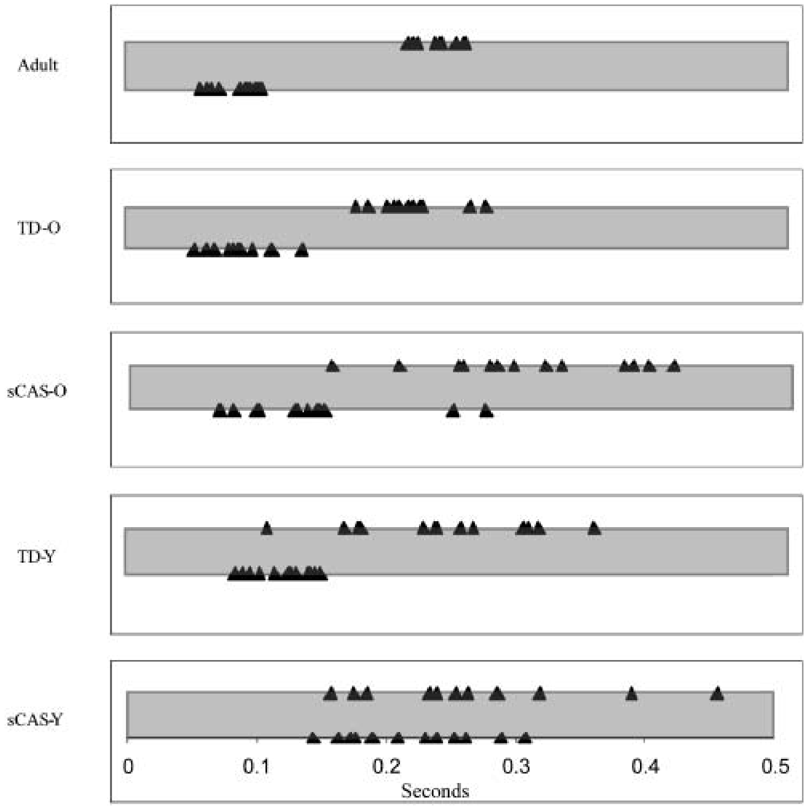


Figure 1. Vowel durations of unstressed (lower line) and stressed (upper line) syllables for each participant during the non-word imitation task.

Closer inspection of the mean durations of the syllable constituents showed that the mean onset durations were remarkably similar across participants, ranging from 0.170 s for TD-O to 0.206 s for sCAS-O. Nucleus durations, by contrast, varied substantially, ranging from 0.238 s to 0.340 s, with greater nucleus durations noted for the two children with sCAS than for their TD peers, whereas systematic variation by age was not evident. Mean coda durations patterned into a decrease with downstep ratios of approximately 1.18 and a rank series of sCAS-Y, TD-Y, sCAS-O, and TD-O, indicating shorter durations for TD status and higher age. Figure 2 shows mean onset, nucleus, and coda durations as a function of participant.

Singing Happy Birthday. Each participant sang the song *Happy Birthday* upon request, and vowel onset intervals were measured, representing the note durations of the song. When each participant's note durations were correlated with those produced by the adult, correlation coefficients systematically distinguished between age and diagnostic status, although the *r* values for the older two participants were relatively close to each other. The highest correlation coefficient, 0.99, was found for TD-O, followed by 0.97 for sCAS-O, 0.94 for TD-Y and 0.74 for sCAS-Y.

The second measure of accuracy, child:adult correlations of $note_x/note_{x+1}$ ratios, yielded a similar distinction, but with more distinct inter-subject differences. Correlation

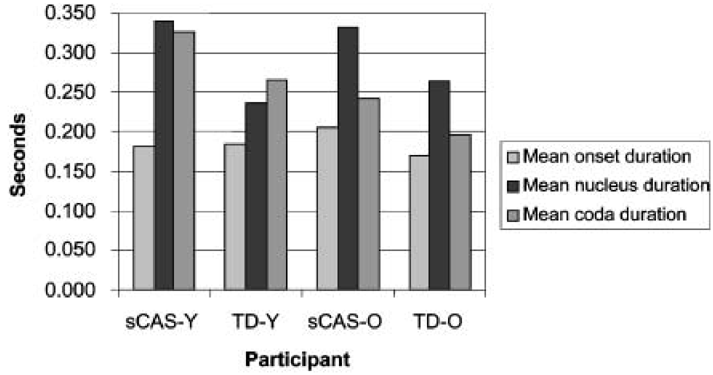


Figure 2. Mean durations of syllable constituents for each participant during the monosyllabic word generation task.

coefficients were as follows: TD-O: 0.97; TD-Y: 0.90; sCAS-O: 0.80; and sCAS-Y: 0.61. In both this, more local, and the preceding, more global accuracy measure, the older participants showed higher rhythmic accuracy than the younger participants, but within each age group, the TD participants outperformed their peers with sCAS.

Regarding the measure of internal rhythmic stability, the relative standard deviation (RSD) for the eight quarter notes in each participant's sample, the lowest RSD, indicating the highest overall stability, was apparent for TD-O (RSD=0.08), followed by sCAS-O (RSD=0.09), sCAS-Y (RSD=0.10), and TD-Y (RSD=0.17). The adult's production was characterized by an RSD of 0.04. This pattern matches the ones from the other two accuracy measures with the exception that the younger TD participant's RSD score was substantially higher than that of her peer with sCAS instead of being lower. All participants inhaled after the word 'you' but TD-Y, who has a history of asthma, inhaled additionally after each occurrence of the word 'birthday', which lengthened the duration of the quarter notes carried by '-day'. Excluding these quarter notes from the sample, the adjusted RSD scores were 0.03, 0.05, 0.07, 0.08 and 0.10 for the adult, TD-O, TD-Y, sCAS-O, and sCAS-Y, respectively, a pattern consistent with an accuracy advantage for TD status and greater age.

Clapped rhythm imitation. Following the first author's model, each participant imitated eight clapped rhythms of varying lengths. Regarding the number of claps produced by each participant in response to the sum of 45 presented claps, the older two participants produced all 45 claps without omissions or insertions. TD-Y produced 37 claps, omitting eight claps and inserting none, which translates into a relative error of 0.18. sCAS-Y omitted two claps and inserted 12, resulting in a relative error (RE) of 0.31.

The subset of claps produced by all participants yielded a datapool of ten intervals, which, when analysed for accuracy, yielded the following child:adult correlation coefficients (r): TD-Y: 0.99; TD-O: 0.98; sCAS-O: 0.96; CAS-Y: 0.92. These results suggest slight accuracy advantages for TD status and higher age.

When the larger sample of intact intervals generated by the two older participants was analysed with respect to accuracy, the child:adult correlation coefficients were 0.97 for TD-O and 0.96 for sCAS-O, indicating a near equivalent proportional accuracy.

Paced repetitive tapping. Participants listened to a metronome beat sequence, then tapped along with the metronome, then continued tapping after the metronome was turned off.

This task was performed at three different metronome speeds. One accuracy measure, indicating how well the participant matched the metronome's overall speed, was a relative error (RE) score, where the absolute difference between the mean tapped interval and the target interval was set proportional to the target interval. Generally, the performance of the TD participants was more accurate than that of the peers with sCAS, and the performance of the older participant within a diagnostic category was more accurate than that of the younger peer. The only exception to this pattern was TD-O at the slowest metronome setting, where her RE score was slightly higher, not lower, than those obtained from both sCAS-O and TD-Y. This score was also higher than TD-O's own score at the other two speeds. When averaged across metronome settings, RE scores were as follows: TD-O: 0.02; sCAS-O: 0.03; TD-Y: 0.07; sCAS-Y: 0.48, a pattern consistent with a TD advantage as well as an age advantage. Figure 3 summarizes RE scores at each metronome speed.

Variability (here an inverse measure of accuracy) was assessed with the relative standard deviation (RSD), where a lower RSD indicates higher accuracy. For each of the three metronome settings, variability profiles showed the same advantage of the TD participant over the peer with sCAS in each age group and an age advantage of the older participant in each diagnostic group over the younger one. When averaged across the three settings, the highest RSD, indicating the greatest amount of variability and the lowest accuracy, was noted for sCAS-Y (RSD=0.16), followed by TD-Y (RSD=0.09), sCAS-O (RSD=0.06) and TD-O (RSD=0.05). Averaged across all participants, accuracy for both the mean tapping interval and variability was greatest in the slowest metronome setting and poorest in the fastest setting. Figure 4 shows RSD scores for all three metronome settings.

Summary of accuracy results. All three music-related tasks and two of the three speech tasks, sentence imitation and non-word imitation, were assessed for accuracy, using a variety of measured units (vowel duration, note duration, interval duration, number of element) and metrics (correlation coefficient, relative error, effect size, ratio, relative standard deviation)

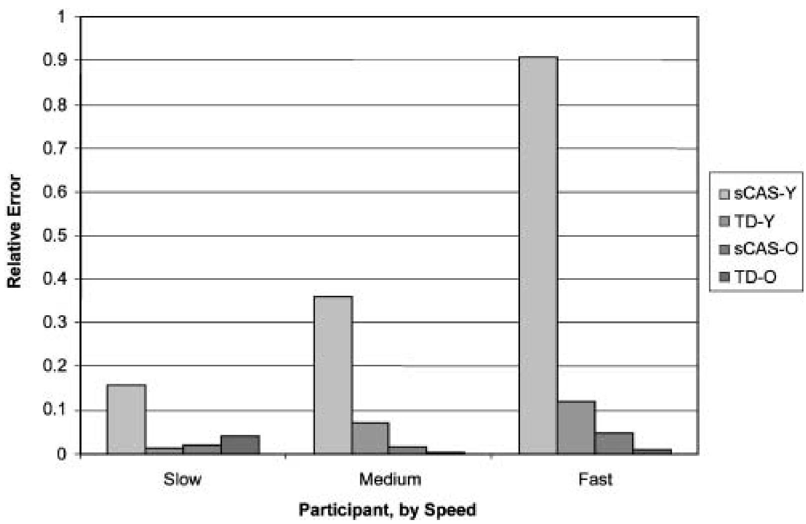


Figure 3. Relative error in speed matching (difference between mean tap interval and target interval proportional to the target interval) during the paced repetitive tapping task.

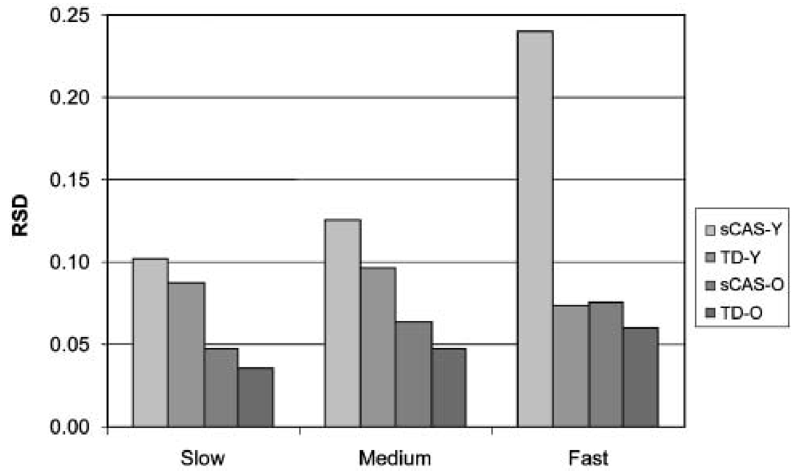


Figure 4. Relative standard deviations (RSDs) by metronome speed and participant during the paced repetitive tapping task.

as indicators, resulting in 11 variables across participants. One speech task, monosyllabic non-word generation, was not analysed in terms of accuracy. Regarding systematic differences within each age-matched pair of participants, ten of the 11 metrics for all speech and music tasks agreed in assigning an accuracy advantage to the TD participant over the participant with sCAS. One metric, relative error of omitted/added claps from the clapped rhythm imitation task, showed absence of errors for both TD-O and sCAS-O. Regarding systematic age-related differences within each diagnostic category, in nine of the 11 metrics the older participant in each diagnostic category outperformed the younger peer, the two exceptions being that TD-Y achieved the same relative error score during the non-word imitation task and a higher child:adult correlation coefficient during the clapped rhythm imitation task, compared to TD-O.

In order to show temporal accuracy as a function of mean unit duration, child:adult correlation coefficients, where available, were averaged across all participants. The task involving the shortest mean duration, 0.181 s, hence requiring the fastest processing and movement, was sentence imitation, and the average child:adult correlation coefficient (r) was 0.64. Non-word imitation, with a mean unit duration of 0.191 s, yielded an average child:adult r of 0.72, followed by clapped rhythm imitation (mean unit duration=0.388 s, r=0.97), and singing *Happy Birthday* (mean unit duration=0.567 s, r=0.82). Figure 5 summarizes child:adult correlation coefficients by task and participant.

Discussion

The purpose of this study was to investigate the timing accuracy in the speech and music productions of two children with suspected childhood apraxia of speech (sCAS) and two age-matched, typically developing (TD) children in order to address two questions, namely (1) are the timing structures in the speech of children with sCAS less accurate and less adult-like than those produced by typically developing (TD) children? and (2) does the degree of temporal accuracy in the speech of children with sCAS have a direct correlate in the domain of music productions?

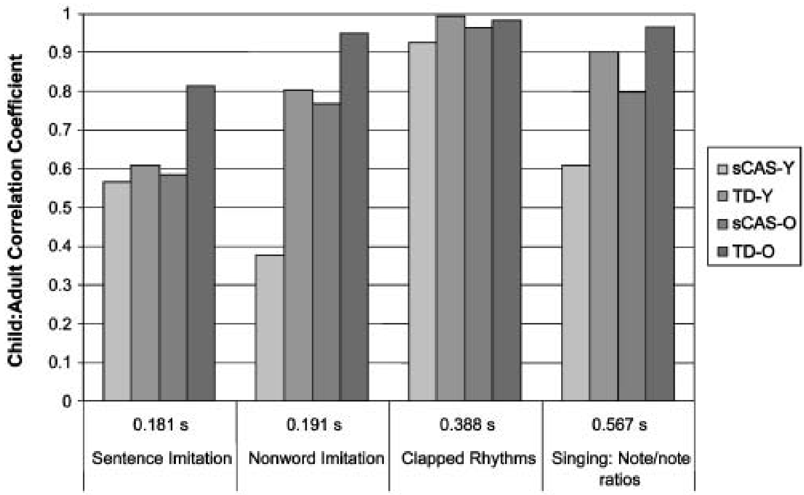


Figure 5. Child:adult correlation coefficients from selected speech and music tasks, by mean unit size.

The first question can be answered with the accuracy profiles obtained from the two speech imitation tasks (sentence imitation and non-word imitation). Metrics for both tasks revealed that the imitations from the two children with sCAS were less accurate (i.e. less similar to the adult model) than those from their TD peers. Measurements for these tasks were based on vowel durations while the non-word imitation task also included an analysis of omitted syllables. This analysis showed that the participants with sCAS produced a less accurate syllable count than their TD peers, omitting weak syllables that did not fit a simple trochaic (i.e. strong-weak) syllable template. Similar results from a non-word imitation task, although obtained with different metrics, have been reported by Rodriguez (1998), who tested three groups of Kindergartners with the same instrument used in the present study, the *Tennessee Test of Rhythm and Intonation* (Koike and Asp, 1981), and found that children with motor speech disorders produced less accurate syllable durations and tended to convert iambic stress patterns (i.e. weak-strong) into trochaic ones. Of all speech tasks in this study, the non-word imitation tasks yielded the largest and most consistent differences between participants with sCAS and their TD peers, which makes it a potential assessment tool of choice in clinical and research settings involving children with sCAS.

Evidence for the answer to the second question, comparing speech and music, comes from the accuracy scores computed from the music-related tasks. During all three tasks and by six out of seven metrics, the participants with sCAS showed lower accuracy than their age peers (in one instance, relative error of hand claps during the clapped rhythm imitation task, both TD-O and sCAS-O did not omit/insert any claps). The most substantial accuracy distinctions between diagnostic categories came from the singing tasks and specifically from the $note_x/note_{x+1}$ ratio correlation. This metric indicates timing accuracy at the local level, pointing to the possibility that the participants with sCAS had greater difficulty than their TD peers in producing a coherent rhythmic structure based on underlying durational relationships. Taken together, these results support the existence of timing deficits during both speech and music-related tasks in the four studied participants and, hence, the existence of a central time keeping mechanism that controls both speech and other time structured activities.

In addition to a consistent accuracy advantage for the TD participant over the peer with sCAS in each pair, a general accuracy advantage for age was observed in all tasks and by all metrics except the clapped rhythm imitation task, where TD-Y achieved a higher child:adult correlation coefficient than TD-O. This finding suggests that timing accuracy, both in oral and limb-generated timing units, increases with age (as would be expected), but it is not clear from the present data what factors drive this development. One possibility is the fact that both of the older participants had music instruction through their school curriculum. Neural development, exposure to time-structured experiences in daily life, such as music, second indicators on watches and clocks, machine noises, awareness of one's own heartbeat, etc., all could promote time keeping accuracy.

It was noted during the paced repetitive tapping task that, averaged across participants, accuracy was highest with the slowest metronome setting (0.577 s) and poorest during the fastest setting (0.375 s). A similar pattern of increased accuracy with increased unit sizes emerged across other tasks, whether speech or music-related. For instance, overall accuracy was relatively high (correlation coefficient range: 0.92; 0.99) during the clapped rhythm imitation task, where the mean clap interval was 0.388 s, and relatively low during sentence imitation (correlation coefficient range: 0.57; 0.81), where the mean vowel duration was 0.181 s, as well as during non-word imitation (correlation coefficient range: 0.15; 0.96), where the mean vowel duration for unstressed syllables was 0.128 s and for stressed, 0.253. While the tasks differed widely in nature and mean durations cannot be compared directly, this trend fits generally with the observation that children's accurate use of reduced syllables is a function of their maturity (Allen and Hawkins, 1980), and with the hypothesis that durational unit sizes may differentially determine processing systems and modes (Gandour, 1987; Zatorre, 2001). The fact that the averaged accuracy was higher during the clapped rhythm imitation task than in the singing *Happy Birthday* task, even though the latter had a slightly higher mean unit duration, may have its explanation in the assumption that the singing task claims more mental and motor resources than the clapping task.

The monosyllabic word generation task was originally added to explore timing differences as a function of syllable structure. According to phonologic theory, lexical stress is strictly assigned on the basis of rime weight, without any contribution from syllable onsets, and nucleus durations are additionally affected by coda voicing and coda obstruency (Gordon, 2002), depending on language-specific characteristics (Buder and Stoel-Gammon, 2002). Regressing a variety of variables such as onset complexity, nucleus complexity, coda complexity, coda voicing, and coda obstruency on each participant's nucleus durations proved problematic because of colinearities in the covariates themselves. When each participant's onsets, nuclei and codas were averaged, however, a comparison of the four profiles showed the following patterns: (1) Mean syllable onsets were of similar duration across participants; (2) mean nucleus durations varied with diagnostic status, where longer nuclei were observed in the children with sCAS; (3) mean coda durations varied with diagnostic status, where longer codas were seen in the participants with sCAS, and (4) mean coda durations also varied with age, where longer codas were seen in the younger participants. The first two findings are consistent with the claims that onsets are irrelevant to stress assignment and, hence, to rime duration (Gordon, 2002), and that vowel nuclei are an excellent predictor of stress distribution (Ramus *et al.*, 2000). Potential interpretations of the longer nucleus durations in the participants with sCAS include an association with slower speaking rate, fewer available programming resources for the upcoming coda consonant, and inadequate acquisition of vowel duration, the latter being a developmental process that has been shown to be robust in young, typically developing children (Kehoe and Stoel-Gammon, 2001; Buder and Stoel-Gammon, 2002). Coda differences on the order of 0.050 s in children as a function of a speech programming disorder and age have not been reported in the literature. One possible interpretation is that syllable-final vocal tract constrictions require higher processing loads than those in onset positions, an interpretation which will have to be balanced against conclusions regarding onset saliency from research in dysfluency, speech errors, and tip-of-the-tongue phenomena. A second possibility is that final lengthening effects on the syllable level are seen more prominently in less mature speakers. In order to understand the present results more fully, a careful analysis of onset and coda characteristics would be necessary.

Future research

A number of questions were not answered with the results from this study. For instance, it remains unclear whether timing inaccuracies in children with sCAS vs. TD children and younger vs. older children are based on perceptual, representational, or production related factors. Also, since this study focused on the timing dimension of prosody exclusively, no observations regarding other acoustic correlates can be offered. Furthermore, other types of childhood speech disorders were not included in the present study, and the TD participants were not matched for gender. The utility of the Pairwise Variability Index (PVI; Low, Grabe and Nolan, 2000) in quantifying the acoustic correlates of any monostress percepts and in distinguishing between speech disorders should be assessed using carefully selected stimulus material. Before claiming timing inaccuracy in speech and music-related tasks as a potential diagnostic marker of sCAS, studies should be carried out with larger numbers of participants, matched not only for age but also for gender; a variety of speech disorders besides sCAS; and tasks that address perception, representation, and production specifically.

Summary

Results from the present study were consistent with previous research in that the participants with sCAS showed greater temporal inaccuracies than their TD peers in their speech. The same profile was observed in their performance of music-related tasks, consistent with the existence of a central time keeping system ('internal metronome') and implying that such a system, not articulatory factors, is the probable source of the observed timing inaccuracies. A corresponding accuracy advantage for the older participants within each diagnostic category was observed in all speech task metrics and six of the seven music task metrics. During one generative speech task, mean onset durations were found to be similar across all participants, while mean nucleus durations differed with diagnostic status and mean coda durations differed both with diagnostic status and age, a result that confirms current theories about stress assignment while also adding new data related to coda durations. Overall, results from this study supported the validity of sCAS as a clinical entity and of temporal fine structure as one of its distinctive dimensions, and confirmed the utility of acoustic analyses.

Acknowledgements

This project was supported by a traineeship grant from the National Institute of Health, NIH 05 T32 DC0003-10. The authors wish to express gratitude to the study participants and their families. We deeply appreciate their gift of time and effort on behalf of communication disorders research. We also thank Chris Moore and Steven Demorest for their valued advice and help in the completion of this project. Portions of this study have been presented at the Child Phonology Conference in Vancouver, BC, 1–4 July 2003, and at the 2003 American Speech, Language, and Hearing Association Convention in Chicago, 13–15 November 2003.

References

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, *75*, 34–46.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child Phonology. Volume 1: Production* (pp. 227–256). New York: Academic Press.

Boersma, P., & Weenink, D. (2003). *Praat Version 4.0.45*. Amsterdam: Institute of Phonetic Sciences.

Buder, E. H., & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America*, *111*, 1854–1864.

Carrow-Woolfolk, E. (1999). *Test for Auditory Comprehension of Language*, third edition. Austin, TX: Pro-Ed.

Davis, B. L., Jakielski, K. J., & Marquardt, T. P. (1998). Developmental apraxia of speech: determiners of differential diagnosis. *Clinical Linguistics and Phonetics*, *12*, 25–45.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Gabrielsson, A. (1999). Music performance. In D. Deutsch (Ed.), *The Psychology of Music*, second edition. San Diego, CA: Academic Press.

Gandour, J. (1987). Tone production in aphasia. In J. Ryalls (Ed.), *Phonetic Approaches to Speech Production in Aphasia and Related Disorders*. Boston, MA: College Hill Press.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation 2*. Circle Pines, MN: American Guidance Service.

Gordon, M. (2002). A phonetically driven account of syllable weight. *Language*, *78*, 51–80.

Green, J. R., Shriberg, L. D., & Campbell, T. F. (2002). Speech timing variables in children with typical speech acquisition, speech delay, and suspected apraxia of speech. Poster presentation at the Conference on Motor Speech, March 2002, Williamsburg, VA.

Guyette, T. W., & Diedrich, W. M. (1981). A critical review of developmental apraxia of speech. In N. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice* (pp. 1–48). New York: Academic Press.

Hayden, D., & Square, P. (1999). *Verbal Motor Production Assessment for Children*. San Antonio, TX: The Psychological Corporation.

Kehoe, M. M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, *28*, 393–432.

Klapp, S. T. (1981). Doing two things at once: the role of temporal compatibility. *Memory and Cognition*, *7*, 375–381.

Koike, K. J., & Asp, C. W. (1981). Tennessee test of rhythm and intonation patterns. *Journal of Speech and Hearing Disorders*, *46*, 81–86.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*, 377–401.

McSweeney, J. L., & Shriberg, L. D. (2001). Clinical research with the prosody-voice screening profile. *Clinical Linguistics and Phonetics*, *15*, 505–528.

Microsoft Corporation. (2002). *Microsoft Excel 2002*. Redmond, WA: Microsoft Corporation.

Morley, M., Court, D., & Miller, H. (1954). Developmental dysarthria. *British Medical Journal*, *1*, 8–10.

Munson, B., Bjorum, E. M., & Windsor, J. (2003). Acoustic and perceptual correlates of stress in non-words produced by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language and Hearing Research*, *46*, 189–202.

Nijland, L., Maassen, B., Van der Meulen, S., Gabreels, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics*, *16*, 461–483.

Peters, M. (1977). Simultaneous performance of two motor activities: the factor of timing. *Neuropsychologica*, *15*, 461–464.

Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693–703.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*, AD3–AD30.

Rodriguez, R. H. (1998). Acoustic and perceptual comparisons of imitative prosody in Kindergartners with and without speech disorders. Unpublished master's thesis. University of Florida.

Schmidt, R. A., & Lee, T. D. (1999). *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL: Human Kinetics.

Semel, E., Wiig, E. H., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals 3*. San Antonio, TX: The Psychological Corporation.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. K. (1997a). Developmental apraxia of speech: I. Descriptive perspectives. *Journal of Speech, Language, and Hearing Research*, *40*, 273–285.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. K. (1997b). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, *40*, 286–312.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. K. (1997c). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, *40*, 313–337.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.

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*, 279–284.

Skinder, A., Strand, E., & 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.

StataCorp. (2001). *Stata Statistical Software: Release 7.0*. College Station, TX: Stata Corporation.

Syntrillium Software Corporation. (1998). *CoolEdit Pro*. Scottsdale, AZ: Syntrillium Software Corporation.

Velleman, S. L., & Shriberg, L. D. (1999). Metrical analysis of the speech of children with developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research*, *42*, 1444–1460.

Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool*. San Antonio, TX: The Psychological Corporation.

Wong, P. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, *59*, 83–95.

Zatorre, R. J. (2001). Neural specializations for tonal processing. *Annals of the New York Academy of Science*, *930*, 193–210.

Appendix A

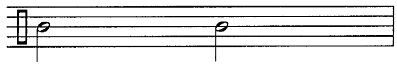
List of stimulus items for the monosyllabic word generation task

A (letter)	fly	ring
bed	grapes	scratch
bee	green	shoe
bread	in	sick
chips	leg	six
cloud	loud	slide
crunch	mad	spin
eat	nuts	splash
egg	oil	sticks
eye	old	string
eyes	pie	tape
fixed	red	

Appendix B

Schematics of rhythms for the clapped rhythm imitation task

1.



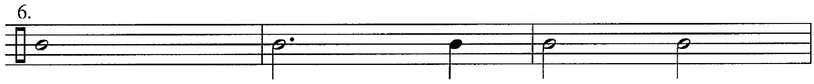
2.

3.

4.

5.

6.



7.

8.