CLASSIFICATION AND MISCLASSIFICATION OF CHILDHOOD APRAXIA OF SPEECH

Phonology Project Technical Report #11

Lawrence D. Shriberg

Jane L. McSweeny

December 2002

Phonology Project, Waisman Center,
University of Wisconsin-Madison

Preparation of this report was supported by research grant DC00496 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (Lawrence D. Shriberg, P.I.)

ABSTRACT[1]

Research in a childhood form of apraxia of speech (CAS) has yet to validate this putative clinical entity and the pathognomonic markers that define it. The present study focuses on speech assessment issues associated with the classification (true positives, true negatives) and misclassification (false positives, false negatives) of CAS. A database of recorded speech samples from nine sources was assembled that included one or more of six types of speech samples from 110 speakers with suspected CAS. One hundred speakers from the five primary sources were classified into eight single, comorbid, and indeterminate etiological categories of speech delay. Classification was based solely on the segmental and suprasegmental information available in the speech sample(s) from each speaker. We describe the development of criteria and quantitative procedures to classify speakers, and report the percentages obtained for each of the eight etiological classification categories. The primary finding is a high rate of possible false positives for CAS. Of the 100 speakers with suspected CAS, approximately 54%–73% met the present study's criteria for one of three alternative single subtypes of speech delay. Secondary findings focus on classification and misclassification issues in relation to the cognitive-linguistic and speech-motor demands of speech tasks used to diagnose CAS. Discussion includes perspectives on design needs in continuing research to identify and validate the behavioral phenotype of CAS.

---

1 Note: The diagnostic marker system developed in this technical report has been modified and expanded in subsequent published studies of acoustic markers of CAS and other proposed etiological subtypes of child speech-sound disorders (Shriberg, Campbell, Karlsson, Brown, McSweeny, & Nadler, in press; Shriberg, Flipsen, Kwiatkowski, & McSweeny, in press; Shriberg, Green, Campbell, McSweeny, & Scheer, in press; Shriberg, Kent, Karlsson, McSweeny, Nadler, & Brown, in press).

Reviews and critiques of research in childhood apraxia of speech (CAS) are available elsewhere (cf. Shriberg, Aram, & Kwiatkowski, 1997a, 1997c; Shriberg, Campbell, et al., in press; Shriberg, Green, et al., in press). The central controversy continues to be the two-fold issues of construct and diagnostic validity: What is the evidence that there is a disorder of praxis that affects the rate and error topology of speech acquisition? What is the necessary and sufficient behavioral marker or set of markers that discriminates this disorder from other etiological subtypes of child speech-sound disorders?

Support for the hypothesis of a developmental (i.e., non-acquired) form of apraxia of speech has been significantly advanced by findings from an extended family studied during the past two decades (e.g., Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998; Hurst, Baraitser, Auger, Graham, & Norell, 1990; Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995; Watkins, Vargha-Khadem, et al., in press). Approximately half of the members of this family have a complex disorder marked by an orofacial apraxia and, reportedly, by an apraxia of speech (cf. Shriberg, Campbell, et al., in press). Specifically, children and adult family members with histories of CAS have speaking patterns that reportedly are similar to the speech patterns of adults with acquired apraxia of speech. Molecular genetic studies have isolated the genotype for the orofacial and speech disorders to a single locus on chromosome 7 (Lai et al., 2000; Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). Neuroimaging and other descriptive analyses continue to refine the neurolinguistic substrates associated with the onset and course of the disorder (e.g., Alcock, Passingham, Watkins, & Vargha-Khadem, 2000; Vargha-Khadem et al., 1998; Watkins, Dronkers, & Vargha-Khadem, in press; Watkins, Gadian, & Vargha-Khadem, 1999). These findings, which will require cross-validation in other molecular genetic studies, are viewed as strong interim support for the validity of CAS as a bonafide clinical subtype of child speech disorder. The eventual availability of a genetic biomarker for CAS will provide the necessary methodological element to address the second need the development of a reliable behavioral phenotype.

Until the availability of both a biomarker and a reliable behavioral phenotype for CAS, it seems likely that a notable recent trend for over-diagnosis of CAS will continue. Davis, Jakielski, & Marquardt (1998), who assessed 20 children referred by speech-language pathologists for suspected CAS, reported that only 5 children (20%) met these authors inclusionary criteria for CAS. A recent survey of 184 caregivers of children with suspected CAS indicated that 20 different classificatory terms are being used to refer to this proposed clinical entity (Lohman, Manning, & Dean, 2001). A claim in one textbook is that the incidence of CAS is 6%–10% (Portwood, 2000), an estimate that greatly exceeds the point-prevalence estimate of 3.8% for undifferentiated speech delay in 6-year-old

children (Shriberg, Tomblin, & McSweeny, 1999).

Likely contributors to over-diagnosis of suspected CAS include information dissemination from a number of disciplinary and advocacy sources. The general message from such sources is that speech-language pathologists are advised to consider CAS as a possible diagnosis for children with late onset of speech, unusual and variable speech errors, and/or protracted normalization despite competent treatment. However, the assessment guidelines for CAS currently are limited to checklists of speech and non-speech characteristics purportedly observed in other children with CAS (cf. Crary, 1993; Davis et al., 1998; Hall, Jordan, & Robin, 1993; Hodge, 1994; McCabe, Rosenthal, & McLeod, 1998; Ozanne, 1995). Thus, the only available diagnostic markers for CAS are those that claim and are perpetuated by a circular form of consensus validity. With both increased administrative press to give each client a diagnostic label in order to receive services, and the general increase in CAS as a possible explanation for severe and persistent child speech-sound disorder, misclassification of CAS is likely to remain at high rates. The goal of this report is to examine some speech assessment issues that bear on classification outcomes.

Speech Tasks

Table 1 is a list of eight types of speech tasks used in protocols to assess children's speech-sound disorders, including children with suspected CAS. The eight tasks are divided into three evocation-content modes, ranging from tasks requiring imitation of nonsense syllables to those evoking samples of spontaneous conversational speech. In addition to the mode of evocation (imitative, non-imitative) and type of content (nonsense and real syllables, words, phrases, sentences, continuous speech), a central dimension of speech samples is their level of complexity or task demand (*challenge*) relative to a speaker's cognitive-linguistic and speech-motor capacities. An investigator's theoretical perspectives on CAS dictate the selection of evocation-content tasks to include in a speech protocol for CAS, the level of complexity needed to evoke proposed markers of CAS, and the quantitative criteria needed to interpret performance as positive or negative for CAS.

A speech assessment protocol that has the *sensitivity* needed to identify a speaker who is a true positive for CAS (i.e., the speaker has the disorder) must include one or more appropriate types of speech tasks at the appropriate level of cognitive-linguistic and speech-motor challenge, regardless of the speaker's severity of involvement. For example, if an investigator uses only a commercially available articulation test to assess performance (an B1 task in Table 1), the test may yield false negatives because the citation word forms may not be sufficiently challenging to evoke markers of CAS in all speakers. Most such citation tasks

consist primarily of common monosyllabic nouns (James, 2001; Morrison & Shriberg, 1992). Similarly, if the investigator uses only a conversational speech sample (C2), behavioral markers of CAS may not occur, because speakers might avoid producing challenging monosyllabic and multisyllabic word forms. As well, if the protocol does not include appropriate tasks and coding procedures to assess prosodic variables such as rate, stress, and timing, it might be insensitive to true positives for CAS if valid pathognomomic markers for CAS reside in these domains.

A speech assessment protocol that has the *specificity* to reject a speaker who is a true negative for CAS (i.e., the speaker does not have the disorder) requires that the tasks evoke behaviors that differentiate typical from atypical speech, as well as markers that identify true positives for alternative disorders (i.e., differential diagnosis). For example, the constructs of *unusual* and *variable* errors are found in the clinical-research checklists used to differentiate CAS from typical speech acquisition, from the most prevalent form of speech delay (i.e., articulation disorder, phonological disorder), and from certain forms of dysarthria. Whether or not a child's pattern of speech errors qualifies as unusual or variable requires standardized reference data appropriate for each of the three comparisons. Unfortunately, standardized reference sources on such constructs as unusual and variable speech errors in typical and atypical speech acquisition are not currently available in the archival literature in child speech disorders. Thus, specificity issues in the possible misclassification of CAS reflect differences in conceptual perspectives on the purview of child speech-sound disorders and significant gaps in available reference data that follow from those perspectives. In addition to these more general considerations, the following discussion focuses on some specific assessment issues associated with false positive and false negative assessment outcomes for children with suspected CAS.

Misclassification of CAS

False positive misclassification of CAS may occur as a consequence of any of three inappropriate conclusions about assessment information. One situation occurs when markers for CAS are not observed during assessment, but an examiner interprets a child's delayed onset of speech or protracted treatment history as sufficient evidence to support suspected CAS. As suggested previously, the typical candidate for such classification is the child who has not made progress comparable to that made by other children who have received a similar type and frequency of treatment. The second situation occurs when speech or other behaviors observed during assessment are incorrectly interpreted as markers of CAS. One such example would be interpreting poor performance on a challenging nonsense word imitation task as evidence of poor speech-motor

planning or sequencing, even though such performance may also or alternatively reflect difficulty in fast mapping associated with deficits in phonological working memory. A third possible source of false positives, as noted previously, may occur when valid markers of CAS observed during assessment are not specific to this disorder. For example, slow speech rate might be viewed as a diagnostic marker for CAS, or alternatively, for dysarthria.

Misclassification of CAS may also occur as false negatives associated with three possible sources. As reviewed above, one possible source occurs when the assessment protocol does not contain speech tasks that are sufficiently challenging to evoke the markers of CAS (i.e., the tasks have insufficient sensitivity). A second possibility is that the speech tasks in a protocol do evoke valid CAS markers, but they are not correctly interpreted or quantified as providing sufficient inclusionary support for CAS. Finally, the tasks used in an assessment protocol and the quantitative criteria for classification may have sufficient sensitivity, but the speaker may have normalized markers of prior CAS available at the level of data reduction used in the protocol (e.g., broad phonetic transcription, narrow phonetic transcription, acoustic-aided transcription, kinematics).

The goals of the present study were to obtain quantitative estimates of sources associated with classification and misclassification of CAS. Owing to limitations in the retrospective design of the study, the current focus is solely on plausible speech measurement issues as they impact classification outcomes.

METHOD

Database

The Phonology Project database was searched for speakers with suspected CAS. This archive of audiocassette and videocassette tapes includes speech samples obtained from local and collaborative projects completed over a span of three decades (cf. http://waisman.wisc.edu/phonology/). The primary inclusionary criterion was that the speaker was suspected to have or have had CAS by a referring speech-language pathologist and/or an investigator. The second criterion was that the sample or samples from each speaker have adequate technical quality and linguistic informativeness for the perceptual methods of this study.

Table 2 includes descriptive information for the assembled database of 108 children and 2 adults with histories of suspected CAS (i.e., the adults had

histories of non-acquired or childhood apraxia of speech). The set of recordings included more than one speech sample for 9 speakers, yielding a total of 121 useable samples, 118 (97.5%) of which had been recorded on audiocassette tapes. Speakers were divided into two groups: *case studies* (sources 1–4) and *small databases* (sources 5–9). The four case studies series (n = 10 speakers) were obtained from correspondence with colleagues and were used to develop the procedures described below. The primary focus of the classification findings to be reported was the five small databases of children originally assessed in local and collaborative projects (n = 100 speakers).

The mean age of speakers in the five databases at the time of assessment was 6 years, 10 months, which is consistent with the generally older ages of children with suspected CAS in research reports (i.e., the children have not made the progress in treatment expected for speakers with typical speech delay). The approximate 7:3 sex ratio favoring boys is consistent with literature findings in both CAS and speech delay, with most of the ratios reported for speech delay based on younger children (Shriberg et al., 1997c; Shriberg et al., 1999). Severity of involvement was indexed by the three metrics shown in Table 2. Each database used comparable speech sampling procedures and the same software for transcription and computerized analysis. The means and standard deviations for each metric were essentially similar across the five databases, with slight differences in means falling within the pooled standard deviations. The average Percentage of Consonants Correct scores across the 100 children (75.3%) placed these generally older speakers in the low mild-moderate range of involvement.

Case studies. As noted above, the 10 speakers comprising the case studies were used to refine the speech and prosody markers for CAS. Sources 1, 2, and 3 were speakers classified as suspected CAS by three clinical researchers with substantial experience in the differential diagnosis of CAS (J. Duffy, personal communication; D. Hammer, personal communication; K. Yoss, personal communication). The three samples in source 3 were the only videotaped samples. The 4 speakers from source 4 are members of the KE family described previously (M. Gopnik, personal communication). Each of these latter speakers, including two children and two adults, had an orofacial apraxia and suspected apraxia of speech.

Small databases. The audiocassette recordings that comprised the small databases were obtained from collaborative research projects. Source 5 included 12 speakers with suspected CAS who were ascertained from the Cleveland area (Shriberg, Aram, & Kwiatkowski, 1997b). Source 6 included 14 speakers with suspected CAS who were referred during the past three decades to a university speech clinic in Madison, Wisconsin. Source 7 included 17 speakers with suspected CAS, as identified by clinical researchers in five cities in North America

(Shriberg et al., 1997c). The 25 speakers in source 8 were recruited primarily from the Pittsburgh area for a genetic study of suspected CAS (Shriberg et al., 2001). Finally, source 9 was a group of 32 speakers primarily from the Cleveland area, who were participating in a genetic study of suspected CAS (Lewis, Freebairn, Hansen, Taylor, & Iyengar, 2001).

Procedures

The authors listened to all of the speech samples in 3–4 hour sessions held 3–4 times per week for a period of approximately 4 months. The videocassette samples were viewed on a Panasonic AG-520B system with a 20-inch monitor, and the audiocassettes were played back on one of several Dictaphone 2025 devices used in prior studies of child speech disorders.

The general procedure during each session was to listen to or view one or more speech samples for each speaker, to compare notes on perceptions, and to reach a consensus on the presence of markers for the etiological classifications of speech disorder to be described. Transcripts of the conversational speech samples, including narrow transcriptions completed by prior research transcribers, were consulted as needed to aid in glossing strings of unintelligible speech. A series of worksheets developed during the early sessions provided formats to annotate, extend, and quantify the speech and prosody markers of CAS proposed in the research literature and described in prior research reports (cf. Odell & Shriberg, 2001). As reviewed previously, the initial listening/viewing sessions focused on samples of children with the highest probability of being true positives for CAS (i.e., the case studies). These sessions placed special emphasis on the behaviors observed in the speakers from source 4, the only speakers to date whose suspected CAS has been associated with a potential genetic biomarker. Throughout all sessions there was no reference to associated case history information or to information from other aspects of assessment protocols, such as performance on non-speech tasks. Thus, all classification decisions were based only on information available in the speech sample(s) for each speaker.

A series of preliminary analyses was completed to resolve procedural issues for the primary analyses. The following subsections describe the issues, findings, and implications of findings for the primary analyses. The Appendix includes text and tabular materials that describe rationale and procedures for each stage of the classification process. Readers will need to first review the terms and classification procedures described in the Appendix before returning to the following text. That is, terms used in the following exposition are defined in the Appendix.

RESULTS AND DISCUSSION

Preliminary Analyses

Speech samples

The goal of the initial preliminary analysis was to inspect the types and complexity levels of the speech samples available from each of the five databases. The most frequent sample types, in order of magnitude, were conversational speech samples (C2; see Table 1), which were available for all (100%) of the 100 speakers, imitated real words in isolation or carrier phrases (A3), available for 71% of the speakers, and non-imitated real words in isolation or carrier phrases (B1), available for 50% of the speakers. Three of the five other speech sample types imitated nonsense words in isolation (A2), imitative sentences (A4), and narratives (C1)--were each available for only 2% of the speakers. The other two types of speech samples, imitated nonsense syllables (A1) and non-imitative phrases or sentences (B2), were not available for any of the present speakers.

There were notable differences in the percentages of speakers in each database for whom the A3 and B1 sample types were available, and the A3 samples from each database differed in complexity of the real word stimuli. Some of the databases had used articulation tests as the A3 stimuli, whereas others had used more complex, phonetically challenging word forms. Figure 1 illustrates the associations of etiological classifications (see Appendix) with speaker age and speech sample type/complexity. The four symbols for the speech samples include simple imitated real words (A3a: circles), challenging imitated real words (A3b: squares), simple non-imitated real words (B1: triangles), and conversational samples (C2: X's).

One observation about the pattern of symbols in Figure 1 concerns the issue of speaker age and the normalization or persistence of markers for disorder. Notice that none of the children meeting criteria for OME (see Appendix for classification terms) in the present study were older than 6 years, suggesting that the speech markers for OME may normalize by this age. In contrast, the ages of speakers classified as SD/RE, DYS, or AOS (AOS is equivalent to CAS in the present context) spanned the ranges shown in Table 2 for the databases. However, consistent with the definition of AOS/DYS (which is used when prosody but not speech markers meet criteria for either AOS or DYS), only 2 children over 6 years of age were classified as AOS/DYS. Associated research indicates that 75% of children with speech delay of unknown origin normalize their deletion and substitution speech errors by 6 years of age (cf. Shriberg et al., 1999). Thus, these data support the suggestion that prosodic markers for CAS and other disorders may persist longer for some children than do speech markers for these disorders (Shriberg et al., 1997c).

A second observation on the pattern of symbols in Figure 1 is that the type and complexity of speech samples used in the listening sessions appears to be associated with classification outcomes. Whereas the OME, DYS, and AOS classifications included one or more speech samples in addition to the conversational sample, most of the AOS/DYS classifications (i.e., indeterminate between AOS and DYS) were made solely from the conversational samples. Thus, for at least some children, more challenging speech tasks (i.e., A3b) may be required to evoke the speech markers that differentiate AOS from DYS.

Marginal/non-marginal ratios

Table 3 includes findings from the second preliminary analysis that focused on marginal versus non-marginal classification outcomes. The summary data for the five small databases (right-most columns) indicate that 56.5% of the classification decisions shown in Table 3 met criteria for marginal classification, and the remaining 43.5% met criteria for non-marginal classification. Three of the databases (sources 6, 7, and 9) had marginal-to-non-marginal ratios that did not depart appreciably from 1:1, but the classifications for speakers from source 5 included somewhat more non-marginal decisions (64.3%), and there were considerably more marginal decisions for speakers from source 8 (81.8%). The subtotaled percentages for each etiological classification (i.e., summed over the five sources) were generally comparable; however, the highest marginal/non-marginal differences occurred for the OME classification decisions (68% marginal, 32% non-marginal). Thus, the five database sources appeared to balance one another, yielding similar proportions of marginal/non-marginal classifications across the single etiological subtypes. These findings are viewed as quantitative support for combining marginal and non-marginal etiological subtype classifications in the primary analysis to follow.

PRIMARY ANALYSES

The primary results of this study are organized by findings and issues for the classification (true positives, true negatives) and misclassification (false positives, false negatives) of CAS. The retrospective design limits the quantitative estimates to the possible false positives, but findings allow for qualitative observations relevant to each of the three other classification outcomes.

False Positives

There were two possible sources of false positives in the present study, based solely on the speech assessment data: the possibility that the original CAS classification resulted from incorrect interpretation of certain speech behaviors as

positive markers for CAS, or the possibility that such verbal behaviors were markers of only CAS. The retrospective design of the present study allows frequency estimates for only the latter. Figure 2 provides these estimates, which will be reviewed after the following discussion of rationale for including or excluding SD/RE classifications in these data.

SD/RE classifications. Rationale for the estimates of false positives in Figure 2 is based on the following considerations. First, 8 speakers classified as NSA were removed from the false positive estimates, yielding a denominator of 92 speakers. In the present data, there were no means to discern why these 8 speakers did not have speech errors consistent with speech delay or residual errors. A hallmark of CAS (again, termed AOS in the classification system described in the Appendix) is the persistence of speech delay. Yet, based on their conversational speech as sampled in the present study, 4 of the 8 children classified as NSA were younger than 6 years of age.

The more difficult procedural question was whether speakers meeting criteria for SD/RE should be included as false positives for AOS. The percentages of speakers classified as SD/RE, with the denominators adjusted for the NSA classifications, ranged from 35.3% to 45.2% across the five databases (Mean = 40.8%: SD = 3.88%). Although these findings support the stability of the classification procedures, it is not clear why approximately 40% of the speakers who were purported to have suspected AOS met this study's classification criteria for SD/RE. These classifications may be considered invalid false positives, because speakers may have normalized their previous AOS markers (but not their other speech errors) at the time the speech samples were obtained. As noted previously, however, the age data indicated that 18 of the 38 children (47.4%) were younger than 6 years of age at the time the speech sample was obtained generally too early for AOS markers to have normalized. Another possible explanation for the large number of SD/RE classifications is that the speech samples were not sufficiently challenging. This explanation is rendered less plausible by the finding that SD/RE classification rates were comparable across databases, despite differences in the types and complexity of speech samples in addition to conversational speech.

An alternative methodological position is that at least some of the SD/RE classifications were valid false positives and should be calculated as such. One observation in support of this perspective is the expectation that conversational samples (available for 100% of the speakers) should be sensitive to the two prosody markers for AOS (see Appendix), which may persist longer than speech markers in children with suspected AOS (Shriberg et al., 1997a, 1997c). Additionally, these 38 children did not have associated fluency or language

involvements behaviors that might confound the classification of AOS versus SD/RE. Recall that the suffixes FI and LI were appended to SD/RE when excessive pauses and other variables affecting stress or timing were thought to be associated with dysfluency or language impairment. Inspection of these data indicated that only 1 of the 38 SD/RE classifications met criteria for the FI suffix, and only 6 classifications met criteria for the LI suffix. Thus, only 7 (18.4%) of the SD/RE classifications could be explained as false positives due to possible interpretations of these speakers' fluency and language involvements as correlates of AOS. For the remaining over 80% of children whose speech was classified as SD/RE in the current study, prior classification as AOS may have been significantly based on information from non-speech assessment tasks and on other historical and treatment outcome information consistent with literature descriptions of childhood apraxia of speech.

These alternative methodological perspectives suggest that the large number of children classified as CAS in the original databases, but as SD/RE in the current study, may be valid or invalid false positive classification outcomes. Therefore, the following estimates are computed both with and without the SD/RE classifications included as false positives.

Findings. The data in the two panels in Figure 2 estimate the percentages of false positives for AOS and the relative frequencies of alternative etiological subtypes as defined in the Appendix. The upper panel, Panel A, assesses what might be termed the *strong* false positive claim, in which speakers classified as anything other than AOS (excepting the indeterminate AOS/DYS) are viewed as false positives. As indicated in the key, the darker bars are these percentages with SD/RE classifications included in the denominators for each calculation (n = 92), and the lighter bars are the percentages with SD/RE classification excluded. With SD/RE classifications included as false positives, false positive rates associated with each of the alternatives to AOS ranged from 1.1% (meeting present criteria for AOS_OME) to 41.2% (meeting study criteria for SD/RE). With SD/RE classifications excluded, false positive rates ranged from 1.9% (AOS_OME) to 20.7% (DYS). Summed across these classification data in the top panel of Figure 2, 64.8% (SD/RE excluded) or 79.4% (SD/RE included) of children who were originally classified as CAS met the present study's misclassification criteria for false positives.

The lower panel in Figure 2, Panel B, provides estimates of false positives based on what might be termed the *weak* claim, in which only the three classifications that do not include AOS (i.e., DYS, OME, DYS_OME) are viewed as false positives (also excepting the indeterminate classification, AOS/DYS). Calculated as such, false positive rates with the SD/RE classifications excluded

and included respectively, were 53.7% and 72.9%.

Discussion. These findings for the percentages of false positives suggest that a primary source of misclassification of CAS may be insensitive and nonspecific assessment relative to other etiological subtypes of speech disorder. Using the methods developed for this study, the four estimates of false positive rates ranged from 53.7% to 79.4%. Thus, approximately 54% to 80% of the samples met classification criteria for one of the five alternative or comorbid proposed etiological subtypes described in the Appendix (i.e., DYS, OME, AOS_DYS, DYS_OME, or AOS_OME. More conservatively, approximately 54% to 73% of the samples met criteria for a classification that did not include AOS (i.e., DYS, OME, or DYS_OME). As indicated by the rank ordering in Figure 2, SD/RE may be the most frequent subtype of speech delay underlying false positives in the clinical-research community, followed by DYS, and then OME.

Clearly, the validity of these estimates of the frequency and sources of false positives in CAS assessment depends on the validity of the markers described in the Appendix for OME, DYS, and AOS, as well as on the validity of the point system criteria. To date, these markers have only limited consensual validity, and the point count system developed for the present needs is otherwise untested. Within these constraints, the distribution of etiological classifications among the five database sources provides some information that may be useful to consider for the design of cross-validation research.

First, inspection of the classification data at the level of databases indicated that each database was associated with proportionally higher classification rates for one of the three primary classification categories, OME, DYS, or AOS (see Table 3). These slight to pronounced weightings toward one of the three disorders could not readily be ascribed to differences in the type and complexity of the available speech samples for each database, because classification percentages differed for databases that had essentially similar arrays of speech samples (see Table 3). Therefore, classification outcomes were more likely associated with local variables in ascertainment and assessment methods.

Another possible source of variance in misclassification of CAS is a speaker's severity of speech involvement, with suspected CAS a more likely classification outcome for children with more severe speech delay. Figure 3 provides severity index information for the 56 speakers who met classification criteria for SD (n = 24; the other 14 SD/RE speakers met criteria for RE), OME (n = 10), DYS (n =16), or AOS (n = 6). To control for differences in ages and gender, the original scores on each of the three metrics in Figure 3 were transformed to *z* scores using the lifespan reference data for typical speech acquisition in Austin & Shriberg (1996). As suggested by the profile for each metric in Figure 3, severity of involvement as assessed from the conversational speech samples was generally not greater for the speakers classified in the present study as CAS. The means

differences and overlapping standard deviations indicated that speakers meeting CAS criteria for the present study actually had somewhat higher average scores on two of the three metrics compared to the speakers meeting criteria for OME and DYS. These data would support the interpretation that etiological subtypes of speech disorder differ on their patterns of speech errors, not on their average severity of expression of the disorder.

To summarize, the findings interpreted as false positives support the prior interpretation of the high proportion of SD/RE classifications in the present study. Specifically, the original classification of each of these 100 children as suspected CAS by the original clinicians/investigators must have been associated with their performance or histories on variables not accounted for in the present methods. Three such possible variables are these children's status on proposed speech markers for CAS not included in the Appendix, their performance on non-speech tasks, and their speech onset and speech treatment histories. Exactly how such information may have contributed to the original classification outcomes of CAS for each child cannot be recovered from the databases with sufficient precision to extend the present analyses.

True Positives

The design of the present study does not allow quantitative estimates of the percentages of true positives, because there presently is no validated phenotype for CAS. However, classification findings suggest some considerations that may be useful for future research designs.

First, as shown in Figure 2, approximately 7% (Panel A) to 22% (Panel B) of the classifications in the present study agreed with the prior classification as CAS, with the exact magnitude of agreement dependent on the tenability of the strong versus weak claim for the definition of a false positive. The research literature typically takes the strong position on this issue, with subject descriptions purporting to report data on children with only CAS (cf. Shriberg et al., 1997b). However, the present findings, summed across the percentages for each comorbid classification (i.e., AOS_OME + AOS_DYS + DYS_OME), indicate that a significant percentage of speakers (9.8% with SD/RE included, 16.8% with SD/RE excluded) meets criteria for these possibly co-occurring disorders. Clinical experience indicates that AOS and DYS frequently are comorbid. The high percentage of false positives associated with DYS (nearly 30% with SD/RE excluded; see Figure 2) may represent the most difficult challenge for differential diagnosis of these two disorders. The previously cited clinical survey of caregivers of children with suspected CAS (Lohman et al., 2001) and others (e.g., Garn-Nunn, 2000) indicate that concurrent involvements of CAS with other developmental disorders are common. Such findings are also supported by the pattern of clinical findings within and across members of the KE family described previously, and are typical

in the description of participants in studies of adult neurogenic speech disorders. Rather than attempting to force participants into single etiological subtypes, which was avoided by the point count system developed for the present study, it would seem prudent to acknowledge and attempt to refine multiple comorbid classification categories for "true" positives for CAS.

A second observation about the relatively small percentage of speakers meeting study criteria for true positives concerns the prosodic domains of stress and timing. Additional analysis indicated that 28 of the speakers whose data are shown in Figure 2 had stress and/or timing patterns that met criteria for the S/T suffix as defined in the Appendix. Thirteen (46.4%) of these speakers were classified as AOS/DYS (i.e., Indeterminate) because they met criteria for stress/timing, but not the speech criteria that would allow classification as AOS, DYS, AOS_DYS, or DYS_OME. Among the children classified as AOS/DYS, 2 met criteria for stress and timing deficits, 10 for stress deficits only, and 1 for timing deficits only. Among the remaining 15 speakers, the S/T suffix was appended to 4 of the 5 (80%) AOS_DYS classifications (3 stress and timing; 1 stress only), 10 of the 19 (52.6%) DYS or DYS_OME classifications (5 stress and timing, 4 stress only, 1 timing only), and 1 of the 6 (16.7%) AOS classifications (stress and timing). Although derived from small samples, these preliminary percentages suggest that (a) the stress/timing deficits described in previous childhood CAS studies (Odell & Shriberg, 2001; Shriberg et al., 1997b, 1997c; Velleman & Shriberg, 1999) are at least, if not more, associated with DYS than they are with AOS, and (b) that stress deficits may be more frequent than timing deficits. As with all perceptual data in the present study, the stability of such preliminary trends for prosodic variables will need to be assessed using instrumental measures of stress (e.g., Shriberg et al., 2001; Skinder, Connahan, Strand, & Betz, 2000; Skinder, Strand, & Mignerey, 1999) and timing (e.g., Green, Shriberg, & Campbell, 2002). There are certainly too few data to speculate on the implications of stress versus timing deficits, respectively, relative to linguistic versus speech motor processing accounts of CAS (cf. Odell & Shriberg, 2001).

True and False Negatives

Quantitative analyses of the sources for true and false negatives in the assessment of children with suspected CAS are also not accessible in the design of the present study, because each speaker was originally classified as positive for CAS. However, findings from the preliminary analyses provide some information on variables that may be associated with true and false classification outcomes.

As reviewed previously, at the time they were sampled, 18 (47.4%) of the 38 children in the present study classified as SD/RE were younger than 6 years of

age, and 20 (52.6%) were older than 6 years. Whether or not SD/RE classifications are viewed as false positives, these findings underscore the possibility that speech markers for CAS may be less available in young children and/or children with limited speech. Moreover, speech markers become less available as children normalize, which requires increasingly challenging speech tasks to evoke a sufficient number of reliable tokens meeting classification criteria for CAS. As well, detection of persisting markers attenuated in frequency and/or articulatory salience may require more sensitive measurement modes. From this perspective, the probabilities of false negatives are increased in both toddlers and school-age children—the former because they have less productive speech to assess, and the latter because normalization processes require more challenging speech tasks to evoke CAS markers and possibly more sensitive data-reduction methods to detect and quantify their occurrence.

From these methodological perspectives, the purportedly increased prevalence rates noted previously for CAS may be, in part, a function of the reduction of false negatives. That is, speech-language pathologists and researchers may more recently be assessing younger children with suspected CAS, using more challenging speech tasks that evoke relevant markers, and possibly using more sensitive and reliable instrumental measurement techniques to detect and quantify potential CAS markers. In conjunction with the increased awareness and the increased number of sources for clinical guidance noted previously, perhaps fewer children with CAS are being misdiagnosed in contemporary compared to previous community screening programs for developmental disorders. From this perspective, the absence of proposed markers, such as delayed speech onset or reduced performance on non-verbal oral motor tasks, may provide veridical support for true negative classifications of CAS.

A third observation relevant to true and false negative classification outcomes for CAS concerns the need for quantitative inclusionary criteria for speech and prosody markers. Recall in the discussion of findings for marginal versus non-marginal classifications that these two levels of inclusionary criteria were approximately equally distributed among the etiological classification categories (Table 3). For the purposes of the primary analyses, the decision was made to collapse this distinction; in the present context, such arbitrary decisions have implications for the possibility of false negative outcomes. If only the non-marginal CAS classifications were considered true positives, the excluded marginal classifications might be false negatives. Such considerations underscore the need for attention to psychometric issues in developing reliable quantitative criteria for CAS speech markers, particularly for screening instruments, in which (by definition) the emphasis is more on sensitivity than specificity.

A final observation relative to true and false negatives addresses the general problem of clinical versus statistical definitions of speech disorder in children. Classification of eight of the present speakers as having normalized speech acquisition (NSA) claims that the speech patterns in the available samples were typical for the speakers' ages. Moreover, the present classification of approximately 38% of the speakers as having SD/RE claims that their speech was not different from that of children with typical speech delay or typical residual errors. Notice that both claims require that such classification be based on appropriate speech samples and appropriate normative reference data and classification criteria. In fact, as in child language disorders, certain child voice disorders, and other domains of communicative disorders, the child speech disorders community has yet to develop a consensus standard on what constitutes the criteria for speech delay at each relevant developmental age. For example, some clinical-research studies include distortion errors in the calculation of severity of involvement, whereas others exclude such behavior or do not differentiate distortion errors from substitution errors in classification schemes (such as the present distinction between SD and RE, which is based on genetic versus environmental sources of variance; cf. Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). In the absence of socially-validated and psychometrically appropriate reference data, the possibility of false negatives for speech delay (with or without associated CAS) remains a potential clinical and research confound. Particularly as early identification programs must consider alternative or comorbid types of speech-language disorders (e.g., late talkers), differentiating true from false negatives for CAS will require more well-developed lifespan data on the boundaries of typical speech and prosody.

CONCLUSIONS

Generalizations from the present findings on classification and misclassification of CAS are constrained by four considerations in the present study: (a) use of retrospective archival sources rather than prospective data; (b) restriction of the available database to speech samples, rather than also including information from non-speech tasks and other case records data; (c) use of auditory-perceptual (i.e., phonetic transcription, prosody-voice coding) measurement, rather than acoustics or other types of instrumental measurement; and (d) use of proposed diagnostic markers and classification procedures that have limited empirical support.

With the above constraints in mind, the primary conclusion from the present findings supports and extends the major perspective expressed in the precedent literature: there appears to be a major problem with the construct validity of CAS as a clinical entity and the features that define it. Even the most conservative estimates of plausible false positives in the present study were unacceptably high. Three alternative etiological origins were proposed for the speech error patterns observed in over 50% of the speakers originally classified as CAS: the common form of speech delay that may be genetically transmitted (SD/RE), dysarthria, and less frequently, the speech residuals of fluctuant hearing loss associated with otitis media with effusion (OME). Findings suggest that such alternative etiological explanations for CAS include the probability of comorbid involvements.

The second conclusion addresses the methodological issues associated with the type and complexity demands of speech assessment tasks in CAS research. In addition to the fundamental need for validated markers of CAS, psychometric issues associated with age of the child, mode of data reduction, and inclusionary criteria are threats to the internal and external validity of classification outcomes. Standard articulation tests and conversational speech sampling may yield false negatives for young speakers who have too little speech for identification by speech or prosody markers. These speech tasks may also yield false negatives for older speakers for whom more challenging task demands may be necessary to evoke criterion markers for CAS. Addressing these methodological needs is one of the many research challenges in continued studies of children with suspected apraxia of speech.

ACKNOWLEDGMENTS

We express our sincere thanks to the following colleagues for their cordial assistance in forwarding the recorded samples used in the present analyses: Dorothy Aram, Thomas Campbell, Barbara Davis, Joseph Duffy, Myrna Gopnik, Penelope Hall, Deborah Hayden, David Hammer, Joan Kwiatkowski, Barbara Lewis, Shelley Velleman, and Kathe Yoss. We also thank Katherina Hauner for excellent editorial assistance with this manuscript. This research was supported by NIH Grant No. DC00496 from the National Institute on Deafness and Other Communication Disorders.

REFERENCES

Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and Language, 75,* 17–33.

Austin, D., & Shriberg, L. D. (1996). *Lifespan reference data for ten measures of articulation competence using the Speech Disorders Classification System (SDCS)* (Tech. Rep. No. 3). Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

Crary, M. A. (1993). *Developmental motor speech disorders.* San Diego, CA: Singular Publishing Group.

Davis, B. L., Jakielski, K. J., & Marquardt, T. P. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics and Phonetics, 12,* 25–46.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature, 18,* 168–170.

Garn-Nunn, P. G. (2000, November). Issues of concern in (suspected) DAS: Listening to families. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, Washington, DC.

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

Hall, P. K., Jordan, L. S., & Robin, D. A. (1993). *Developmental apraxia of speech: Theory and clinical practice.* Austin, TX: Pro-Ed.

Hodge, M. M. (1994). Assessment of children with developmental apraxia of speech: A rationale. *Clinics in Communication Disorders, 4,* 91–101.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology, 32,* 352–355.

James, D. G. H. (2001). An item analysis of Australian English words for an articulation and phonological test for children aged 2 to 7 years. *Clinical Linguistics and Phonetics, 15,* 457–485.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Levy, E. R., Hodgson, S., Fox, M., Jeremiah, S., Povey, S., Jamison, D. C., Green, E. D., Vargha-Khadem, F., & Monaco, A. P. (2000). The SPCH1 region on human 7q31: Genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *American Journal of Human Genetics, 67,* 357–368.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413,* 519–523.

Lewis, B., Freebairn, L., Hansen, A., Taylor, H. G., & Iyengar, S. (2001, June). *Family pedigrees of children with suspected developmental apraxia of speech.* Poster presented at the 22nd Annual Symposium on Research in Child Language Disorders, Madison, WI.

Lohman, P. S., Manning, R. K., & Dean, C. R. (2001, November). Current trends in developmental apraxia of speech. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

McCabe, P., Rosenthal, J. B., & McLeod, S. (1998). Features of developmental dyspraxia in the general speech-impaired population? *Clinical Linguistics and Phonetics, 12,* 105–126.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35*, 259–273.

Odell, K. H., & Shriberg, L. D. (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics and Phonetics, 15,* 275–307.

Ozanne, A. (1995). The search for developmental verbal dyspraxia. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 91–109). San Diego, CA: Singular Publishing Group, Inc.

Portwood, M. (2000). *Understanding developmental dyspraxia.* London: David Fulton Publishers.

Shriberg, L. D. (2002, November). *Classification and misclassification of child speech sound disorders*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Atlanta, GA.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36, 105–140.*

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (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. (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. (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., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40,* 708–722.

Shriberg, L. D., Campbell, T. F., Brown, R. L., Karlsson, H. B., McSweeny, J. L., & Nadler, C. J. (2001, November). *Toward an acoustic marker for childhood apraxia of speech.* Poster presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., & Nadler, C. J. (in press). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Special Issue: Diagnostic Markers for Child Speech-Sound Disorders, Clinical Linguistics & Phonetics.*

Shriberg, L. D., Flipsen, P., Jr., Kwiatkowski, J., & McSweeny, J. L. (in press). A diagnostic marker for speech delay associated with otitis media with effusion: The intelligibility-speech gap. *Special Issue: Diagnostic Markers for Child Speech-Sound Disorders, Clinical Linguistics & Phonetics.*

Shriberg, L. D., Green, J. R., Campbell, T. F., McSweeny, J. L., & Scheer, A. (in press). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Special Issue: Diagnostic Markers for Child Speech-Sound Disorders, Clinical Linguistics & Phonetics.*

Shriberg, L. D., Kent, R. D., Karlsson, H. B., McSweeny, J. L., Nadler, C. J., & Brown, R. L. (in press). A diagnostic marker for speech delay associated with otitis media with effusion: Backing of obstruents. *Special Issue: Diagnostic Markers for Child Speech-Sound Disorders, Clinical Linguistics & Phonetics.*

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1461–1481.

Skinder, A., Connahan, K., Strand, E., & Betz, S. (2000). Acoustic correlates of perceived lexical stress errors in children with developmental apraxia of speech. *Journal of Medical Speech Pathology, 8,* 279–284.

Skinder, A., Strand, E. A., & Mignerey, M. (1999). Perceptual and acoustic analysis of the lexical and sentential stress in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 7,* 133–144.

Vargha-Khadem, F.,Watkins, K. E., Alcock, K. J., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 930–933.

Vargha-Khadem, F.,Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., Frackowiak, R. S., Friston, K. J., Pembrey, M. E., Mishkin, M., Gadian, D. G., & Passingham, R. E. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 12695–12700.

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,* 1444–1460.

Watkins, K. E., Dronkers, N., & Vargha-Khadem, F. (in press). Behavioural analysis of an inherited speech and language disorder: Comparison with an acquired aphasia. *Brain.*

Watkins, K. E., Gadian, D. G., & Vargha-Khadem, F. (1999). Functional and structural brain abnormalities associated with a genetic disorder of speech and language. *American Journal of Human Genetics, 65,* 1215–1221.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., Frackowiak, R. S. J., Mishkin, M., & Gadian, D. G. (in press). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain*.

APPENDIX: CLASSIFICATION PROCEDURES

Worksheets

The process of etiological classification began by completing worksheets that tallied each of the marker behaviors in Table A1 as they occurred in a speech sample. These behavioral markers for the three proposed subtypes of speech delay of currently unknown origin listed in Table A2 (abbreviated to SD-AOS, SD-DYS, and SD-OME) were assembled from prior and ongoing research in etiologic classification of child speech-sound disorders (e.g., Shriberg, 2002). Brackets were placed around markers on the worksheets if there was only borderline support (generally fewer than three occurrences) for their presence in a speech sample. Additional reasons for uncertainty included (a) technical constraints on the quality of the tape recording, (b) a very mild form of the behavior, and/or (c) lack of a clear consensus between the two listeners. In contrast, non-bracketed markers on the worksheets indicated that support for their occurrence was considered non-ambiguous and sufficient.

Marginal and Non-marginal Classification

The second stage of the classification process was to develop quantitative criteria to translate marginal and non-marginal worksheet data into marginal and non-marginal etiological classifications. Table A1 includes the simple point system used to derive point counts for each classification. Trials of alternative systems indicated that a 2-point system appeared to maximize the dual objectives of validity and reliability, using the information on the worksheets for each speaker. The resulting quantitative classifications matched our overall qualitative classification decisions for each speaker. The system was highly reliable, requiring only clerical vigilance to derive a sum from the tallies of the bracketed and non-bracketed markers on the work sheets.

Multiple and Indeterminate Involvements

A third need in the classification process was to refine a series of decision criteria for speakers for whom a single classification assignment (i.e., AOS, DYS, OME, or SD) was not appropriate. The following procedural conventions were developed to accommodate three such situations.

Multiple involvements. Some speakers met point-count criteria for more than one of the proposed etiological categories. Multiple classification categories were developed for such situations, including AOS_DYS (i.e., AOS *and* DYS),

AOS_OME, and DYS_OME. Three-way classifications (i.e., AOS_DYS_OME) were not obtained for any speaker.

Stress and timing. Preliminary analyses indicated that some children met the stress or timing criteria for AOS or DYS, but not the speech criteria that might differentiate the two classifications. Sufficient support for these behaviors in continuous speech qualified a speaker for the speech or timing (i.e., S/T) suffix (i.e., AOS_S/T, DYS_S/T, or AOS_DYS_S/T). For those speakers meeting criteria for only the S/T marker, a classification termed AOS/DYS (i.e., AOS *and/or* DYS) was developed. This classification indicated that such speakers were indeterminate relative to AOS or DYS using the procedures in this study.

Subtypes of Speech Delay. The classification framework was also modified to accommodate subtypes within Speech Delay (SD), which was the default classification in the procedure to be described. One need was to expand the SD category to include children with residual errors, which for the present purposes could be treated as the residuals of speech delay (i.e., speech delay or residual errors, SD/RE). Another need was to account for the possibility of fluency deficits in association with SD, which could underlie observed differences in stress and timing. A suffix convention was developed to indicate speech delay or residual errors and *fluency involvement* (i.e., SD/RE_FI). A final need was to indicate which children also had language involvement, correlates of which could also underlie observed differences in stress and timing. The suffix LI was used to denote children with speech delay or residual errors and language involvement (i.e., SD/RE_LI).

Classification Procedures

Table A2 describes the procedures to classify speakers into one of the eight disorder classes using the point count criteria: SD/RE, OME, DYS_OME, DYS, AOS_OME, AOS_DYS, AOS/DYS, and AOS. Children not meeting criteria for one of these classifications were assigned Normal or Normalized Speech Acquisition (NSA), indicating no speech delay or residual errors evident in any one or more of the speech samples.

Table A1. *Speech and prosody markers and bracketing criteria for three of the four single etiological classification categories.*

Classification	Behavior	Point Values	
		Bracketed	Non-Bracketed
SD-OME	1. Epenthetic vowels on glide consonants	.25	.5
	2. Frequent glottal stops in all positions	.25	.5
	3. Glottal stops before word-initial vowels	.25	.5
	4. Backing of fricatives and/or stops	.25	.5
	5. Nasal interchanges/confusion	.25	.5
	6. /h/-initial deletions; /h/ insertions on word-initial vowels; /h/-initial substitutions	.5	1.0
	7. Initial consonant deletions	.5	1.0
	8. Correct liquids, especially in clusters	.5	1.0
SD-DYS	1. Excessive/equal stress (Shriberg et al., 1997b)	0.0	0.0
	2. Inappropriate timing (syllable segregation; isochrony) (Green et al., 2002)	0.0	0.0
	3. Slow rate	.5	1.0
	4. Voice tremor/tremulous	.5	1.0
	5. Nasopharyngeal resonance	.5	1.0
	6. Nasal resonance	.5	1.0
	7. Nasal emission	1.0	2.0
	8. Imprecise vowels/consonants	1.0	2.0
SD-AOS	1. Excessive/equal stress	0.0	0.0
	2. Inappropriate timing: (syllable segregation; isochrony)	0.0	0.0
	3. Inconsistent errors on the same word type	.5	1
	4. Prearticulatory oral gestures (groping)	1.0	2.0
	5. Postarticulatory repetitions and revisions	1.0	2.0
	6. Metathetic/sequencing errors	1.0	2.0

Table A2. *Definitions and criteria used to assign speakers to one of the eight etiological classification categories.*

Classification Category	Abbreviation	Description	Criteria and Symbolization
Speech Delay-Apraxia of Speech	AOS	Speech Delay in association with apraxia of speech	Speakers meet phonological process or residual distortion error criteria for SD/RE, plus have a total of 1–1.5 points (marginal) or 2 points or greater (non-marginal) for the eight segmental and prosody markers for AOS described in Table A1. Denote speakers who also meet criteria for either of the two stress/timing markers by the suffix S/T (i.e., AOS_S/T).
Speech Delay-Apraxia of Speech or Dysarthria	AOS/DYS	Speech Delay in association with apraxia of speech and/or dysarthria	Speakers meet only the marginal or non-marginal stress and/or timing criteria for AOS or DYS in Table A1 (i.e., they do not meet the segmental criteria for either or both of these classifications).
Speech Delay-Apraxia and Dysarthria	AOS_DYS	Speech Delay in association with apraxia of speech and dysarthria	Speakers meet criteria for both AOS and DYS. Place brackets around stems meeting only marginal criteria for that classification (i.e., AOS_[DYS]). Denote speakers who also meet criteria for either of the two stress/timing markers for AOS or DYS by the suffix S/T (i.e., AOS_DYS_S/T).
Speech Delay-Apraxia and Otitis Media with Effusion	AOS_OME	Speech Delay in association with apraxia of speech and otitis media with effusion	Speakers meet criteria for both AOS and OME. Place brackets around stems meeting only marginal criteria for that classification (i.e., [AOS]_OME). Denote speakers who also meet criteria for either of the two stress/timing markers for AOS by the suffix S/T (i.e., AOS_OME_S/T).
Speech Delay-Dysarthria	DYS	Speech Delay in association with dysarthria	Speakers meet phonological process or residual distortion error criteria for SD/RE, plus have a total of 1–1.5 points (marginal) or 2 points or greater (non-marginal) for the 6 segmental and prosody markers for DYS described in Table A1. Denote speakers who also meet criteria for either of the two stress/timing markers for DYS by the suffix S/T (i.e., DYS_S/T).
Speech Delay-Dysarthria and Otitis Media with Effusion	DYS_OME	Speech Delay in association with dysarthria and otitis media with effusion	Speakers meet criteria for both DYS and OME. Place brackets around stems meeting only marginal criteria for that classification (i.e., [DYS]_OME). Denote speakers who also meet criteria for either of the two stress/timing markers for DYS by the suffix S/T (i.e., AOS_S/T).
Speech Delay-Otitis Media with Effusion	OME	Speech Delay in association with OME	Speakers meet phonological process or residual distortion error criteria for SD/RE, plus have a total of 1–1.5 points (marginal) or 2 points or greater (non-marginal) for the eight segmental markers for OME described in Table A1.

Speech Delay/Residual Errors	SD/RE	Speech Delay or residual distortion errors	Speakers' segmental errors are consistent with natural phonological processes (SD) or typical residual distortion errors (RE), as defined in Shriberg (1993; Appendix). Denote speakers who also meet marginal or non-marginal worksheet criteria for fluency involvement (i.e., repetitions, revisions, blocks, and/or prolongations) by the suffix FI (i.e., SD/RE_LI). Denote speakers who also meet marginal or non-marginal worksheet criteria for language involvement (i.e., errors in morphosyntax, telegraphic speech, word retrieval, language formulation, and/or low MLU) by the suffix LI (i.e., SD/RE_FI).
Normal or Normalized Speech Acquisition	NSA	Typical speech and prosody	Speakers whose speech and prosody do not meet criteria for any of the eight single or multiple etiological classifications described above.

Table 1. Eight types of speech tasks used in protocols to assess Childhood Apraxia of Speech (CAS). Within each task the stimuli reflect a gradient from *simple* to *challenging,* with challenging tasks posing greater cognitive-linguistic and speech-motor demands.

A. Imitative responses to auditory models, pictures, and/or text
1. Nonsense syllables in isolation or carrier phrase
2. Nonsense words in isolation or carrier phrase
3. Real words in isolation or carrier phrase
4. Phrases and sentences
B. Non-imitative responses to pictures or text
1. Real words in isolation or carrier phrase
2. Phrases and sentences
C. Spontaneous Continuous speech
1. Narratives
2. Conversation

Table 2. *Description of the 110 speakers with suspected childhood apraxia of speech (CAS).*

			Sex			Age (yrs;mos)[a]			Speech					
									PCC[b]		PVC[b]		II[b]	
Sources		*n*	*n* Male	*n* Female	% Male	M	SD	Range	M	SD	M	SD	M	SD
Case Studies														
	Source 1[c]	1	1	0	100	14;10	–	–	–	–	–	–	–	–
	Source 2[c]	2	2	0	100	7;11	–	6;0-9;10	–	–	–	–	–	–
	Source 3[c]	3	3	0	100	3;5	0;6	3;0-4;0	–	–	–	–	–	–
	Source 4	4	2	2	50	–	–	4;0-43;0	80.8	9.4	89.7	4.6	89.8	9.3
	Subtotal/Mean	10	8	2	80	–	–	3;0-43;0	–	–	–	–	–	–
Databases														
	Source 5	12	11	1	92	8;11	3;4	5;9-14;11	70.6	15.5	84.1	6.3	83.7	16.9
	Source 6[d]	14	11	3	79	5;3	1;3	3;5-8;4	70.5	9.7	94.1	3.3	84.6	12.3
	Source 7	17	7	10	41	8;3	3;5	4;8-14;4	74.7	14.5	94.3	4.4	89.8	8.3
	Source 8	25	18	7	72	6;10	2;5	3;3-12;0	83.9	11.8	94.3	3.7	90.9	9.2
	Source 9	32	24	8	75	6;2	2;3	3;0-13;6	72.5	12.3	94.6	2.9	83.6	12.0
	Subtotal	100	71	29	71	6;10	2;9	3;0-14;11	75.3	13.5	93.1	5.1	86.7	11.8
	Grand Total/ Mean	110	79	31	71.8	–	–	3;0-43;0	–	–	–	–	–	–

[a]Some ages from sources 2, 3, and 4 were estimated.

[b]PCC: Percentage of Consonants Correct; PVC: Percentage of Vowels Correct; II: Intelligibility Index (Shriberg, Austin, et al., 1997).

[c]Continuous speech samples were not available for speakers from sources 1, 2, and 3.

[d]Speech measures were available for 12 of the 14 samples.

Table 3. *Distribution of marginal (M: <2 points) and non-marginal (NM: ≥ 2 points) classifications among the five database sources*[a].

		AOS/DYS				OME				DYS				AOS				All			
		M		NM		M		NM		M		NM		M		NM		M		NM	
Databases	n[b]	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Source 5	7	0	0	0	0	1	100	0	0	0.5	12.5	3.5	87.5	1	50	1	50	2.5	35.7	4.5	64.3
Source 6	8	2	40	3	60	0.5	100	0	0	1.5	60	1	40	0	0	0	0	4	50	4	50
Source 7	11	1	25	3	75	2	100	0	0	3	60	2	40	0	0	0	0	6	54.5	5	45.5
Source 8	11	1	100	0	0	1.5	100	0	0	2	66.7	1	33.3	4.5	81.8	1	18.2	9	81.8	2	18.2
Source 9	17	3	100	0	0	3	42.9	4	57.1	3	54.5	2.5	45.5	0	0	1.5	100	9	52.1	8	47.1
Totals/ Mean	54	7	53.8	6	46.2	8	66.7	4	33.3	10	50	10	50	5.5	61.1	3.5	38.9	30.5	56.5	23.5	43.5

[a]Comorbid classifications were tallied by assuming 0.5 for each single classification (i.e., AOS_OME: 0.5 for AOS and 0.5 for OME).

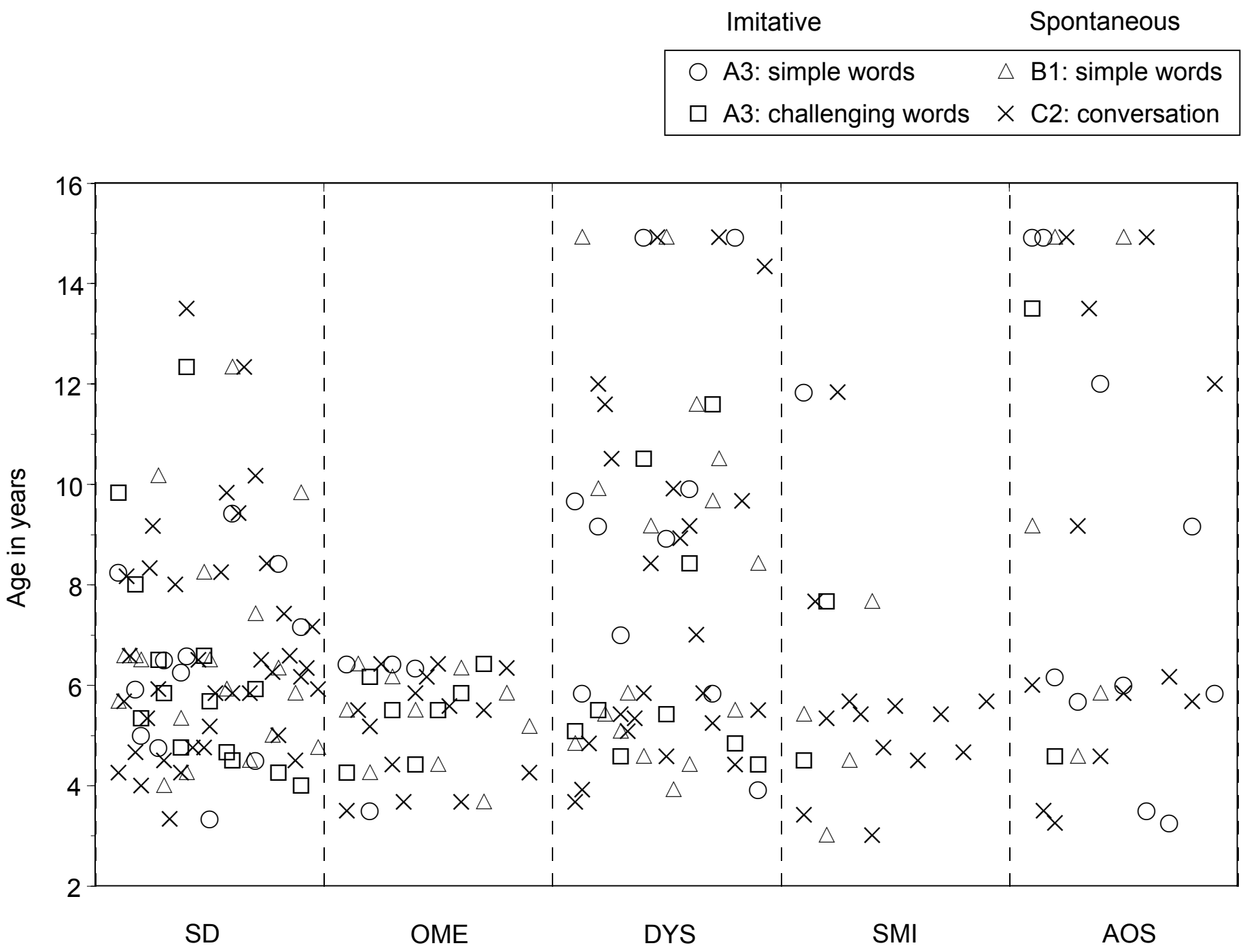
[b]Adjusted for number of NSA and SD/RE classifications.

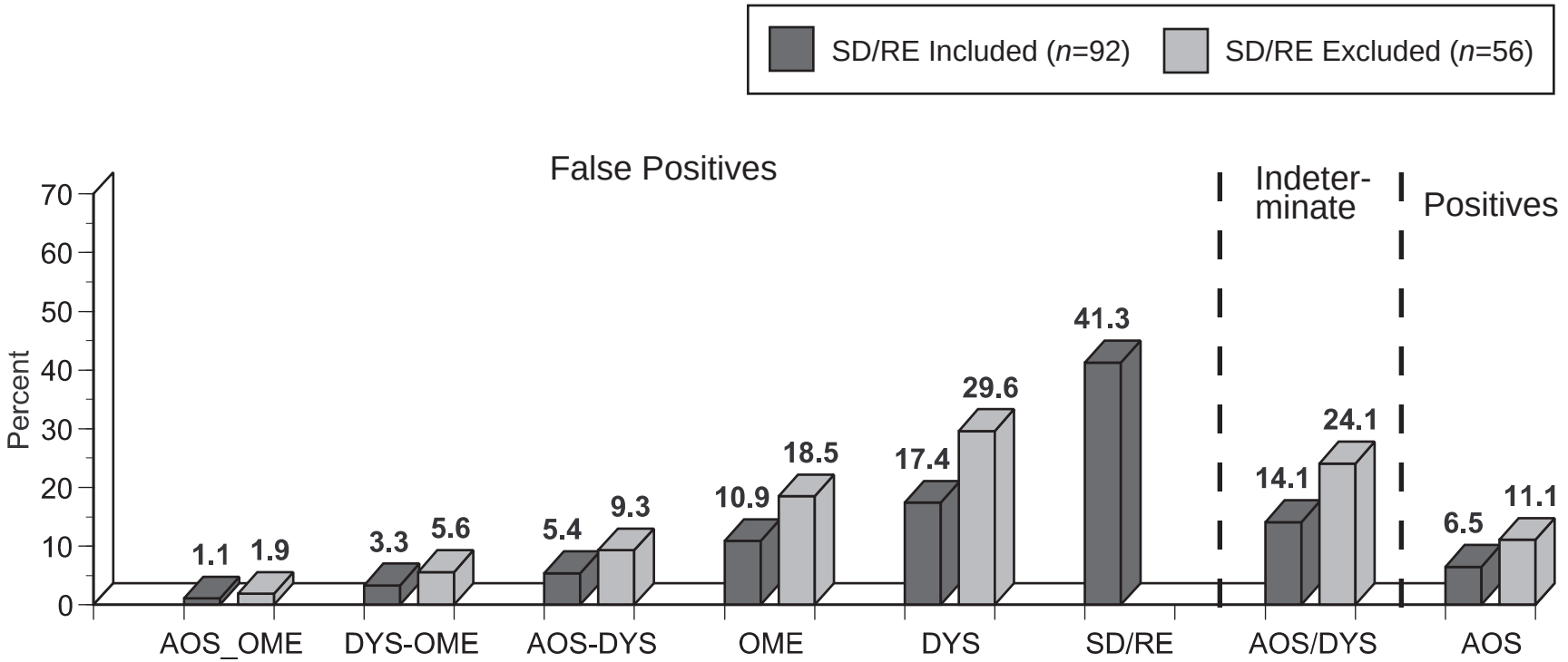
Figure Captions

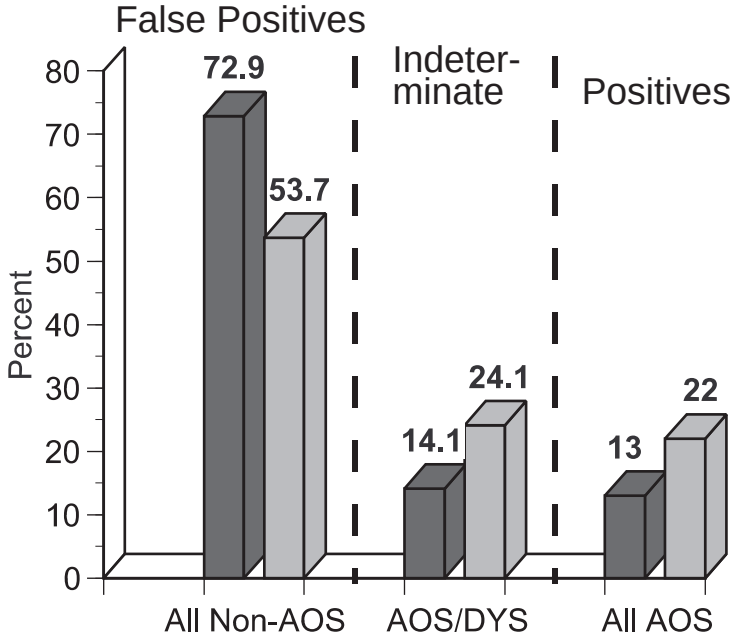
Figure 1. Associations among speaker age, type and complexity of speech sample, and etiological classification outcomes.

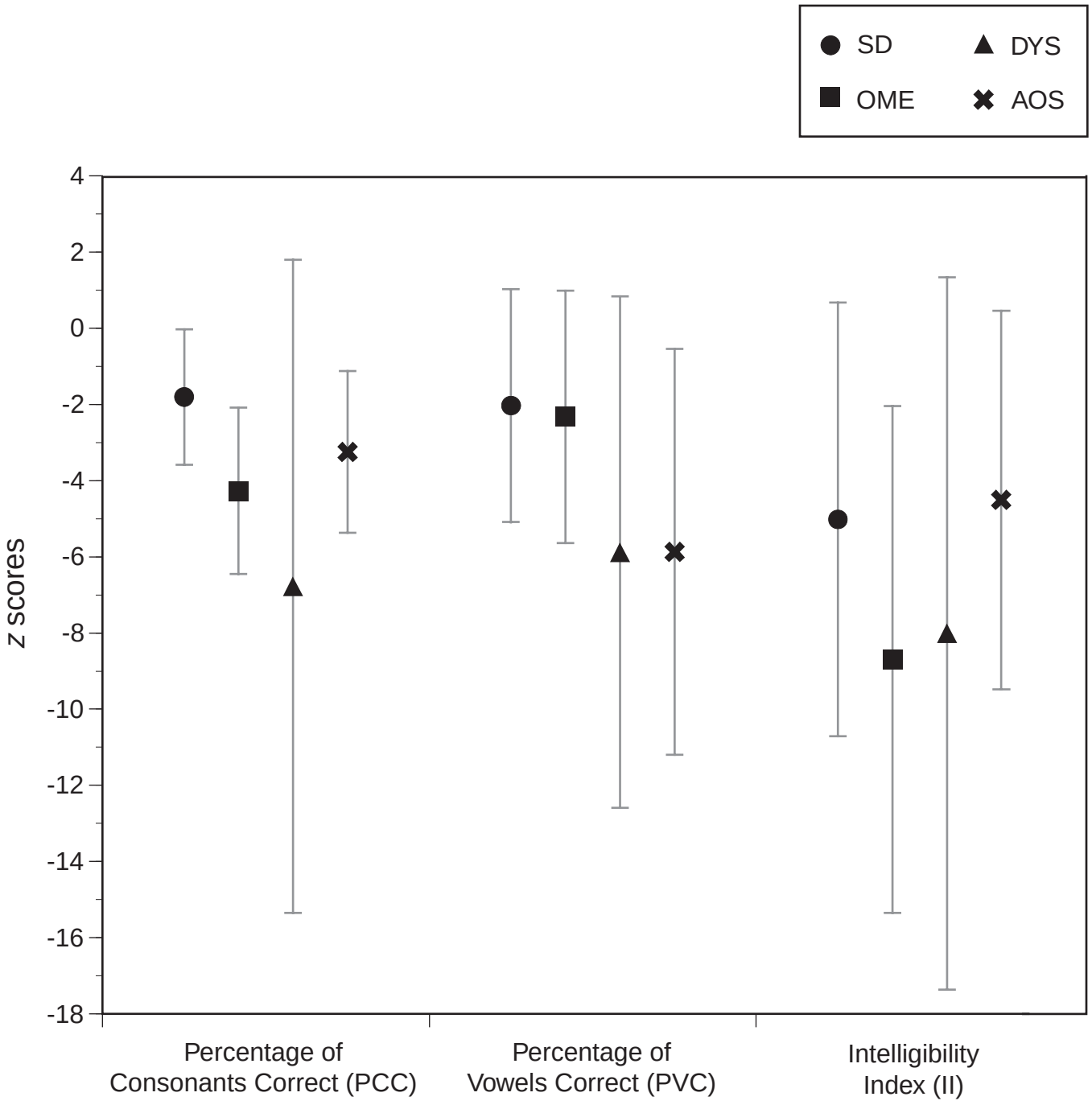
Figure 2. Estimates of the percentages and sources of false positives in 100 children originally classified as suspected CAS. The etiological classifications were based on speech sample findings using procedures described in the Appendix. See text for assumptions underlying the data in the upper versus lower panels.

Figure 3. Speech severity comparisons for the 56 speakers who met classification criteria for SD (24), OME (10), DYS (16), or AOS. The *z* scores for each of the three severity metrics (see Table 2) were computed from the lifespan reference data for these measures reported in Austin & Shriberg (1996).


Imitative
Spontaneous
A3: simple words
B1: simple words
A3: challenging words
C2: conversation
Age in years
16
14
12
10
8
6
4
2
SD
OME
DYS
SMI
AOS


SD/RE Included (n=92)
SD/RE Excluded (n=56)
False Positives
Indeter-minate
Positives
Percent
70
60
50
40
30
20
10
0
1.1
1.9
3.3
5.6
5.4
9.3
10.9
18.5
17.4
29.6
41.3
14.1
24.1
6.5
11.1
AOS_OME
DYS-OME
AOS-DYS
OME
DYS
SD/RE
AOS/DYS
AOS


False Positives
Indeter-minate
Positives
Percent
80
70
60
50
40
30
20
10
0
72.9
53.7
14.1
24.1
13
22
All Non-AOS
AOS/DYS
All AOS


SD
DYS
OME
AOS
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
z scores
Percentage of Consonants Correct (PCC)
Percentage of Vowels Correct (PVC)
Intelligibility Index (II)