# Crosslinguistic Application of English-Centric Rhythm Descriptors in Motor Speech Disorders

**Julie M. Liss**, **Rene Utianski**, and **Kaitlin Lansford**
Department of Speech and Hearing Science, Arizona State University, Tempe, Ariz., USA

## Abstract

**Background—**Rhythmic disturbances are a hallmark of motor speech disorders, in which the motor control deficits interfere with the outward flow of speech and by extension speech understanding. As the functions of rhythm are language-specific, breakdowns in rhythm should have *language-specific* consequences for communication.

**Objective—**The goals of this paper are to (i) provide a review of the cognitive- linguistic role of rhythm in speech perception in a general sense and crosslinguistically; (ii) present new results of lexical segmentation challenges posed by different types of dysarthria in American English, and (iii) offer a framework for crosslinguistic considerations for speech rhythm disturbances in the diagnosis and treatment of communication disorders associated with motor speech disorders.

**Summary—**This review presents theoretical and empirical reasons for considering speech rhythm as a critical component of communication deficits in motor speech disorders, and addresses the need for crosslinguistic research to explore language-universal versus language-specific aspects of motor speech disorders.

## Keywords

Acoustic analysis; Dysarthria; Perceptual analysis; Speech intelligibility; Speech perception

> *Constrained by the phonetic structures of the languages of the world, speech rhythm is the production, for a listener, of a regular recurrence of waxing and waning prominence profiles across syllable chains over time, and with the communicative function of making speech understanding in various speaking styles more effective. [1, p. 41].*

Rhythmic disturbances are a hallmark of motor speech disorders, in which the motor control deficits interfere with the outward flow of speech. In fact, these disturbances are so perceptually salient that the vast majority of the classic perceptual symptoms of Darley et al. [2] refer to the various aspects of rhythmic breakdown in motor speech disorders: reduced stress, monopitch, monoloudness, slow rate, short phrases, increase of rate in segments, increase of rate overall, variable rate, prolonged intervals, inappropriate silences, short rushes of speech, excess and equal stress, prolonged phonemes, repeated phonemes,

Julie M. Liss, Department of Speech and Hearing Science, Arizona State University, Box 870102, Tempe, AZ 85287-0102 (USA), julie.liss @ asu.edu.

irregular articulatory breakdown, and distorted vowels. The gold standard for differential diagnosis of motor speech disorders involves identifying constellations of these perceptual symptoms, as they are the behavioral consequences of the underlying neurological disease [2, 3].

This etiology-based view carries with it an important assumption of *universality*: just as people around the world with Parkinson's disease are identifiable by their shuffles, stoops and tremors, so too should their rushed, mumbled speech be iconic to the trained speech-language pathologist. That is, the perceptual symptoms of rhythmic breakdown should be identifiable, irrespective of the language being spoken, because the symptoms are the manifestation of the underlying *movement* disorder. One can gain an appreciation of the common acceptance of this etiology-based view by examining the frequency and range with which the Mayo Classification System is referenced globally. Table 1 contains the results of a literature search in PubMed, Medline, and Google Scholar, from 1969 to present for all non-American English articles on dysarthria that reference the Mayo system in participant speech characterization [search items: ref. 2–4 or ref. 5, diagnosis, classification]. Although this search was by no means exhaustive, it yielded 51 peer-reviewed publications, spanning 23 languages and dialects, thereby supporting the notion of a global acceptance (for the list of articles, see online supplementary 'Appendix A', www.karger.com/doi/10.1159/000350030).

Despite their widespread application, there is good reason to question the suitability of English-centric descriptors, particularly rhythm descriptors, for characterizing the *communication disorder* caused by motor speech disorders crosslinguistically. Speech rhythm serves important linguistic functions to facilitate speech processing, and these linguistic functions are language-specific. Thus breakdowns in rhythm due to motor speech disorders should have *language-specific* consequences for communication. In an earlier paper on speech rhythm in the dysarthrias we made the following observation [6, p. 1346]:

> We have suggested and cited evidence that this reduction in temporal contrast is a source of intelligibility decrement for English listeners who rely on this cue for lexical segmentation [e.g., Liss et al., 2000]. But would this be the case to the same extent for listeners less inclined to rely on this cue because of its lack of relevance in their own language, such as Spanish or French? It is conceivable that the rhythm abnormalities in dysarthria – and perhaps other aspects of speech deficit as well – cause fundamentally different challenges for listeners across languages.

The implications of this are not lost on those who study and treat motor speech disorders in non-English languages. Indeed, Ma et al. [7] and Whitehill [8] have published a number of papers that explicitly address language-specific and language-universal effects of various dysarthrias in English versus the tone languages of Mandarin and Cantonese. The dysarthria associated with Parkinson's disease has been investigated in a range of languages, wherein it has been reported that the symptom of reduced fundamental frequency variation in speech ('monopitch') has a language-universal effect on speech prosody, and a language-specific effect for those languages in which fundamental frequency serves a phonological function [7, 9–11]. Chakraborty et al. [12, p. 268], who conducted a perceptual analysis of dysarthria in the language of Bengali, acknowledged, '… since speech sounds (phonetics) and patterns

of stress and intonation in speech (prosody) appear to vary in the context of individual languages, findings from studies on dysarthric speech in other languages cannot be applied to Bengali speech'. It is a problem in search of an answer.

We take the opportunity in this paper to: (i) review the cognitive-linguistic role of rhythm in speech perception in a general sense and crosslinguistically; (ii) present novel results of lexical segmentation and degraded rhythm studies in American English dysarthrias, and (iii) offer a framework for crosslinguistic considerations for speech rhythm disturbances in the diagnosis and treatment of communication disorders associated with motor speech disorders.

## The Cognitive-Linguistic Role of Rhythm in Speech Perception

'Speech is inherently tied to time' [13, p. 392]. Common to languages is the propensity for the speech stream to be unpacked as a wave of acoustic-perceptual prominences and troughs. As Kohler [1] explained in his historical account of the study of rhythm in speech and language, the field developed out of perceptual experiences that languages have qualitatively different rhythms. These perceptual impressions (e.g., Spanish as 'machine gun', and English as 'Morse code') were operationalized as the hypothetical language rhythm classes of syllable-timed and stress-timed, respectively [14]. According to Kohler [1], the field eventually moved away from its perceptual foundations and toward a quantitative segment-duration focus. The goal of this work was to establish and then populate these hypothetical rhythm classes from the sorting of crosslinguistic durational metrics, without consideration of rhythm perception. Despite intense effort, the approach of rhythm-as-timing has not yielded a rhythm-based language classification scheme. Some of the reasons for this failure are methodological, in that the duration-based rhythm metrics are susceptible to extralanguage influences[1], which complicate or preclude their interpretation [see, for example, ref. 17–19]. Yet other reasons for the failure of this classification scheme can be linked to underlying assumptions about the perception of speech rhythm. For example, the approach requires that rhythm classes are perceived categorically. White et al. [20] recently presented evidence that this is not the case. Their English-speaking listeners attended to a variety of temporal cues in graded ways – including speaking rate, vowel and consonant durations, and utterance-final lengthening – to discriminate among languages and dialects. Yet, perhaps the most critical source of failure is the assumption that rhythm perception lies exclusively in the temporal domain [e.g., 21–23].

This ignores the potential contribution of complementary spectral regularities in the speech signal [see ref. 1]. As described by Arvaniti [18, p. 351]:

> Traditional descriptions of speech rhythm have relied on the notion of isochrony,
> that is, the idea that rhythm rests on regulating the duration of particular units in
> speech, syllables in syllable timed languages, stress feet in stress-timed languages,

---

[1]This is the result of several factors. First, speech segmentation for these analyses is based on phonological operational definitions – vocalic and intervocalic segmentation at the phoneme level – which carry assumptions about the role of these units in rhythm perception [1, 15]. These assumptions remain to be demonstrated as valid within and between languages, and there is already strong evidence that crosslinguistic validity is questionable [16]. Second, temporal patterns are closely tied to the composition of the speech material such that it is not warranted to make direct comparisons across studies using different stimuli [17]. Third, results vary depending on the particular type of rhythm metrics employed [e.g., 16], as some are speaking rate-normalized, some are raw values, some are ratio metrics, and some are more susceptible to speech material and individual speaker variation than others [17].

and moras in moratimed languages. Thus in this view, rhythm is based exclusively in durational patterns or timing (indeed the terms rhythm and timing have often been used as synonyms in this literature […]).

The call for a conceptual shift back to rhythm *perception* and to its *communicative function* is growing louder [see ref. 1, 19, 24–27]. Such a focus is necessary for the development of a crosslinguistic model of the role of speech rhythm disturbance in communication disorders. Specifically, this shift will permit exploration of how communicative function of speech rhythm in a given language predicts the impact of degradation of those rhythm cues.

We begin by adopting Kohler's [1] broad definition that the perception of speech rhythm is subserved by recurring spectral-temporal patterns of change in fundamental frequency, spectral amplitude, syllable duration, and spectral dynamics. Further, the usefulness of these cues – in isolation or bundled – depends on a variety of factors, including the rhythmic structure and phonology of the language and dialect being spoken; the native language and dialect of the listener, and the details of the communicative interaction [25, 28]. The key concept in this definition is that the universal communicative function of rhythm is to *reduce the computational load of speech processing*. Specifically, rhythmic patterns allow the listener to *track*, *segment*, *anticipate*, and *focus attention* on high yield aspects of the speech signal [1, 29, 30] and to facilitate recognition and prediction of syntactic and semantic relationships among words and phrases in continuous speech [31, 32]. Thus, our working definition of speech rhythm for the purposes of this review is as follows:

- recurring spectral-temporal patterns

- which occur locally (syllable, word) and are distributed (clause, sentence, discourse) in an interdependent hierarchical framework from sounds to meaning,

- which are perceptible and communicatively meaningful to listeners,

- and which universally function to reduce the computational load of speech processing[2].

Importantly, we also conceptualize *rhythm* as a superordinate of other related linguistic constructs such as prosody, prominence, stress, strength, markedness, etc. This view is consistent with a rapidly growing body of neurophysiological evidence for the existence of endogenous, dynamic neural oscillators that drive rhythmic expectancy and anticipatory responses in speech perception and production [e.g., 32, 34–41]. Our innate sensitivity to rhythmicity permits an information-processing framework that can integrate across modalities (auditory, visual, somatosensory), time scales (speech sounds, words, phrases, discourse, facial expression, body language, gestures, etc.), and intercommunicator details (e.g., turn-taking, emotional state, conversational entrainment, etc.). While an in-depth treatment of the neurophysiological literature is beyond the scope of this review, the theoretical implications emerging from this work are critical for conceptualizing the *communicative* functions of speech rhythm crosslinguistically, and particularly for

---

[2]Langus et al. [31], using an artificial language and Italian listeners, convincingly showed the hierarchical structure of prosodic cues in speech segmentation at multiple levels. This work, as well as previous work by Cummins [33], demonstrate how tracking speech rhythm reduces the computational load of speech processing.

conceptualizing communication disorders as they manifest on the continuum from diminished speech intelligibility to disrupted conversational interactions [40, 42, 43].

Whereas the basic functions of speech rhythm in speech perception seem to be universal features of human spoken communication, their application is surely language- specific as noted in Kohler's [1] definition. Cumming [25, p. 22] cautions,

> … it should not be assumed that native speakers of all languages perceive rhythm identically; investigating rhythm perception from a crosslinguistic perspective is essential if a universal view of the phenomenon is desired which is not biased by the weighting of cues in any particular language.

This observation has strong support in findings that the rhythmic structure of a listener's native language biases the ways in which he/she uses rhythm cues for various speech perception tasks [e.g., 44–48]. For example, Huang and Johnson [49] reported that American English and Mandarin Chinese speakers used acoustic cues for intonation in Mandarin differently. Whereas Mandarin speakers responded to tone contour with sensitivity for phonologically relevant contrasts, English speakers responded to pitch level. Interestingly, this difference in cue use persisted even when the phonological information was eliminated from the speech stimuli. This is not surprising: language-specific rhythm properties are among the most robust cues exploited by infants in speech and language development, and as such, enjoy a lifetime of reinforced use and refinement [e.g., 50–52].

## Rhythm and Lexical Boundary Identification

As an example of language-universal and language-specific uses of speech rhythm, we turn our attention to the cognitive-perceptual process of lexical segmentation of connected speech. Regardless of a language's rhythmic structure, the first task of a language learner is to parse the acoustic stream into the words that comprise it. Accurate parsing is essential for the subsequent development of vocabulary and grammar. Much research has shown that statistical probabilities drive decision-making whether the language learner is an infant being exposed to a native language [e.g., 53] or an adult learning an artificial language [e.g., 54]. Two sources of probability data appear to be flexibly used depending on their relative informativeness: speech rhythm (as defined herein) and phonotactic probabilities. The important feature of this learning is that the cue preference is influenced by the rhythmic and phonotactic structure of the language being heard [e.g., 47, 48]. McQueen and Cutler [55, p. 508] made the observation that '… rhythm allows a single, universally valid description of otherwise very different segmentation procedures used across languages'. Table 2 contains an example of rhythmic cue use for lexical segmentation in several languages, for the purposes of illustration.

In English, rhythmic cues for lexical segmentation are found in the increased durations of segments associated with the edges and heads of syntactic domains: word-final lengthening [e.g., 56], phrase-final lengthening [e.g., 57], stressed syllable lengthening [e.g., 58], accentual lengthening (i.e. the increased duration of segments in words that carry phrasal stress/pitch accent) [e.g., 59]. Syllabic rhythm cues also are found in the spectral domain,

with strong syllables distinguished from weak syllables by vowel formant structure and peaks in fundamental frequency and amplitude contours [60, 61].

Because the probabilities of the English language favor stressed syllables as word onsets, listeners rely on the perceptual prominence of strong syllables relative to weak syllables for lexical segmentation. Rhythmic cues of import appear to be those that signal syllabic stress in English, including variation in fundamental frequency, syllable/vowel duration, and vowel quality or strength [e.g., 62, 63]. This has been described by Cutler and Norris [64] as *the metrical segmentation strategy* (MSS). Supporting evidence for this is in the lexical boundary error (LBE) patterns that favor insertion of lexical boundaries prior to strong syllables [65, 66].

Mattys et al. [67] proposed a hierarchical model of cues to lexical segmentation that has been an important framework for conceptualizing intelligibility deficits in the dysarthrias. In this model, listeners call on various knowledge sources to guide lexical segmentation depending on the quality and quantity of acoustic and contextual information available, defaulting to the most efficient and economical solutions. In highly contextualized, good-quality speech, lexical segmentation occurs as word recognition. That is, the listener perceives the speech stream as a string of words and no explicit segmentation is required. As the speech signal and complementary cues become impoverished, listeners must resort progressively to more active lexical segmentation by using remaining acoustic cues. When distortion precludes reliable phoneme identification in connected speech, listeners shift their attention to *rhythmic cues* to make weighted predictions about word onsets and offsets [see ref. 65, 67–69]. Segmental ambiguities are then resolved within these word-sized frames. Thus, the role of rhythm in speech segmentation is elevated under adverse listening conditions [see ref. 70, for a valuable overview of speech recognition in adverse conditions].

The reliance on perceptual prominence cues in the face of phonemic uncertainty works well when these cues are intact, as these cues tend to be robust and discernible even in noise [28]. But when the rhythm cues themselves are reduced or abnormal, this cognitive-perceptual strategy is challenged. The relationship between the type of prosodic degradation (i.e. which traditional cues to segmentation are diminished) and the resulting perceptual errors is extremely useful for explaining *why* the intelligibility deficit occurs.

## Lexical Segmentation and Degraded Rhythm in American English Dysarthrias

We have established thus far that speech rhythm, as defined herein, serves important language-specific communicative functions. We also have established that speech rhythm cues are especially important for the communicative function of lexical segmentation when the speech signal is degraded. The next step is to ask whether we can explain (and eventually predict) how the nature of the rhythm degradation relates to a particular pattern of perceptual consequences. This explicitly requires us to consider how listeners use speech rhythm cues as clues to word boundaries, and, as has been our theme throughout, the cues and clues are language-specific. In this section of our review, we present data that demonstrate how degradation of the speech rhythm cues in American English interferes with

the process of lexical segmentation. The goal of our work has been to develop a predictive model that accounts for the relationships among speech deficit patterns[3] and their perceptual consequences. This has involved relatively large-scale projects, which have generated a corpus of speech samples and listener transcriptions of experimental phrases designed for conducting lexical segmentation analysis. A number of published studies have established the viability of the LBE paradigm in dysarthric speech [68, 69, 71, 72].

In the conception of this line of research, we identified four dysarthria subtypes whose speech disturbance patterns had relatively little overlap, in accordance with the gold standard classification system [4]. We recruited speakers with neurological disease who exhibited characteristic perceptual features in their speech. People with hypokinetic dysarthria secondary to Parkinson's disease presented with rushed and mumbled articulation, a hypophonic voice, and little pitch variation. Those with hyperkinetic dysarthria secondary to Huntington's disease displayed unpredictable bursts of speech, along with unusual articulation and nonspeech vocalizations. People with ataxic dysarthria secondary to cerebellar disease presented with an intoxicated speech quality, irregular in its breakdown and timing; and finally, those with a mixed flaccid-spastic dysarthria secondary to amyotrophic lateral sclerosis spoke very slowly, prolonging syllables toward isochrony, but maintaining pitch variation. We reasoned that, despite the inherent messiness of naturally degraded speech, the import of 'prosodic cues' (as we referred to them then, now subsumed under the more global term 'rhythm') to lexical segmentation of dysarthric speech should be revealed in the resulting LBE patterns. This is, by and large, what we found [68, 69, 72]. Specific findings and observations from our previously published reports are summarized in table 3.

## New Data

This brings us to the new data resulting from our most comprehensive analysis of lexical segmentation of dysarthric speech, in which the four dysarthria subtypes were directly compared (see 'Appendix' for full methodological details). We examined LBE patterns elicited from 44 speakers with dysarthria. These speakers were selected on the basis of cardinal perceptual characteristics associated with hypokinetic, hyperkinetic, ataxic and mixed flaccid-spastic dysarthria subtypes [4], as summarized above, and all speakers presented with at least moderate severity. The data for this experiment consisted of transcriptions of phrases spoken by individuals with dysarthria, transcribed by 60 healthy young listeners. The 80 six-syllable phrases were specifically designed to permit LBE analysis and interpretation, such that they were of low semantic predictability and followed either iambic or trochaic stress patterning [65].

The null hypothesis for this analysis was as follows: If the form of the speech degradation is inconsequential to the task of lexical segmentation, we would expect the patterns of LBEs to be similar across transcriptions of dysarthria subtypes. The listener transcripts were coded independently by two trained judges to assess the nature of lexical segmentation errors.

---

[3]Although speakers with dysarthria were selected because their speech characteristics corresponded with those of the Mayo Classification System diagnosis, we are now evaluating a much broader range of dysarthria presentations, irrespective of etiology, to develop this predictive model.

Importantly for this report, the judges first identified errors in lexical segmentation, and then designated the type of error (either insertion, I, or deletion, D) and the location of the error (either before a strong, S, or a weak, W, syllable). For example, hearing 'I can't' for the target 'attend' is an erroneous lexical boundary insertion before the strong syllable 'tend' (IS), and hearing 'sewer' for the target 'sue her' is an erroneous lexical boundary deletion before a weak syllable, 'her' (DW). This permitted the LBE pattern analysis relative to the outcomes hypothesized by the MSS hypothesis, which predicts that listeners tend to erroneously insert lexical boundaries before strong (IS) rather than weak syllables (IW) and erroneously delete boundaries before weak (DW) than before strong syllables (DS). This prediction was evaluated for four dysarthria subtypes in which the rhythmic structure differed and therefore was expected to yield different LBE patterns.

Before turning to the LBE pattern results, it is first necessary to assess the comparability of the LBE pools in terms of equivalent magnitude of speech severity across the four groups of speakers with dysarthria. Attaining comparable pools allows us to consider LBE pattern differences relative to the dysarthria subtypes, rather than to overall severity, for example. Table 4 shows the percent of words correctly transcribed for the four speaker groups, along with the total number of LBEs for each group. Visual inspection of these data reveals that three of the groups were of highly similar intelligibility and this was confirmed by statistical analysis. As shown in table 5, while results of the analyses of variances evaluating the effects of dysarthria group on the percent of words correct [$F_{(3, 56)} = 17.969$; $p < 0.001$, $\eta^2 = 0.490$] and the number of LBEs [$F_{(3, 56)} = 4.880$; $p < 0.01$, $\eta^2 = 0.207$] were significant, pairwise comparisons showed that hyperkinetic speech was of significantly lower intelligibility than that of the ataxic, hypokinetic, and mixed groups, which did not differ from each other in any pairwise comparison. The larger number of errors for the hyperkinetic group also was accompanied by a greater total number of LBEs[4]. Based on these results, we can safely compare and interpret LBE patterns among these latter three groups, and we must consider the possibility of severity effects when interpreting the LBE patterns for the hyperkinetic group relative to the others.

Figure 1 offers a visual representation of the LBE data across the four dysarthria groups. Each bar is the ratio of *predicted* to *nonpredicted* error types (IS/IW; DW/DS), based on the MSS hypothesis. The front row of bars corresponds with predicted insertion errors (IS/IW) and the back row shows predicted deletion errors (DW/DS). A ratio of '1' indicates equal proportions of predicted and nonpredicted errors, and higher values indicate a stronger adherence to the MSS predicted error patterns. It should be pointed out here that the corpus of phrases was designed to contain slightly more opportunities for non-predicted than predicted errors. This means, then, that a ratio above 1.0 can be interpreted with a fair degree of confidence of conformity with the MSS hypothesis.

The first question is whether or not listeners used available cues to treat strong syllables as word onsets within each group. The answer is yes: $\chi^2$ tests of independence, conducted within each dysarthria group (d.f. = 1), demonstrated dependence between LBE type

---

[4]This is not always the case, as LBEs are only one type of error. Other errors include word substitutions that do not violate a lexical boundary (such as 'flashing' for 'smashing'), or errors of word omission or nonresponses.

(insertion or deletion) and location (before strong and before weak). The results of the $\chi^2$ tests of independence (reported in table 6) were all significant at $p < 0.001$. The LBE patterns elicited from each of the dysarthrias conformed with the MSS hypothesis, such that predicted LBEs outnumbered nonpredicted errors, albeit to different extents (see table 4 for proportions and distributions of LBE types, for each dysarthria group). Specifically, for each speaker group, the percent of total LBEs designated as deletions before weak syllables (DW) exceeded that of deletions before strong syllables (DS), and proportion of insertions before strong syllables (IS) exceeded that of insertions before weak syllables (IW). Thus, listeners attended to the available rhythmic cues to syllable stress to guide lexical segmentation, and tended to assign strong syllables as word onsets.

This finding is expected, as these native American English listeners should use the strategies specific to the English language in their attempts to understand degraded speech. However, the second and more important question is whether or not the various forms of degradation posed different challenges to these strategies. Figure 1 and the associated $\chi^2$ goodness of fit analyses reported in table 6 confirm this to be the case. We compared the LBE distributions between groups to test the null hypothesis that they were drawn from the same sample. Significant findings were interpreted as indicating that the distributions differed between the dysarthria groups. Results indicated that *none* of the four dysarthria distribution patterns was drawn from the same distribution, as all results were significant at $p < 0.001$. That is, the patterns of LBEs, even though all adhering to the MSS hypothesis (as confirmed by the results of the $\chi^2$ test of independence), were significantly different from one another. So, even within languages, rhythmic differences elicit patterns that conform to the MSS, but in different degrees; thus, it is a logical expectation that languages with different rhythmic structures will produce some language-specific results that must be included in a theory of the communication deficit in dysarthria.

To address how the pools differed, let us turn first to our three comparable groups with equivalent intelligibility and total numbers of LBEs. The hypokinetic pool had reduced predicted-to-nonpredicted ratios, with the highest proportion number of lexical boundary deletion errors of the three groups. The mixed pool had the highest proportion of insertion errors, and the highest IS/IW ratios. The ataxic pool was intermediate in error distribution[5]. Because dysarthria severity as assessed by overall speech intelligibility cannot account for the differences, we look to the distinctive ways in which the speech rhythm is disturbed in the groups, based on perceptual assessment and acoustic measures conducted on these phrases in our previous work.

Table 7 contains a comparison of the current LBE findings with our previous rhythm metric findings [6] as an entry point for interpretation. Hypokinetic speech was distinctive in its elicitation of the highest proportion of lexical boundary deletion errors for any group. This tendency for listeners to 'compress' adjacent syllables may be linked to the rapid speaking

---

[5]It should be noted that these findings for ataxic dysarthria LBE patterns are quite different from those reported in Liss et al. [68]. Those results were based on a different group of speakers with ataxic dysarthria who presented with largely equal and even stress. The present group, by perceptual and acoustic metrics, did not exhibit this characteristic, rather they produced slow speech with irregular articulatory breakdown. This discrepancy high-lights the need to move toward acoustic-perceptual characterization rather than diagnostic classification for studies of communication.

rate, monotonicity, and imprecise articulation, all which converge to blur word boundaries. Yet, listeners both erroneously deleted and inserted lexical boundaries most often in the predicted locations. This is explained by the rhythm metrics that showed relatively preserved temporal relationships among vocalic and intervocalic intervals. Thus, despite being reduced, listeners were able to use the available contrast cues to guide lexical segmentation.

The mixed flaccid-spastic dysarthria was at the other end of the acoustic spectrum, with very slow speech that was syllabified such that all vowels were drawn out, lacking temporal distinctiveness. This LBE pool was distinctive in the large proportion of LBE insertions. Just as a propensity to erroneously delete lexical boundaries in rushed speech, the slowed and syllabified speech promoted erroneous insertions. However, the insertion errors were predominantly before strong syllables, indicating that listeners had access to syllabic strength information. We speculate that intonation contours served as one important cue for this, as the mixed group had relatively preserved fundamental frequency variation.

The most distinctive aspect of the ataxic LBE pool was the relative normalcy of the error proportions and adherence to the MSS. Indeed, this pool had the strongest $\chi^2$ test of independence results, underscoring the dependence of error type and location in this distribution. In other words, there were equal amounts of deletions and insertions before both strong and weak syllables (see table 6). Rhythm metrics conducted on these phrases showed vocalic and intervocalic temporal contrasts with values overlapping those of healthy control speakers. Even though intelligibility was reduced to the same degree as hypokinetic and mixed, rhythmic contrast was largely preserved.

Whereas the hyperkinetic pool cannot be compared directly to those of the other three dysarthrias because the phrases were less intelligible, we can address the LBE findings relative to the perceptual-acoustic characterization of the phrases. The hyperkinetic phrases can be characterized perceptually as emerging as unpredictable fits and starts, with occasional loud bursts, random fluctuations in pitch, all of which culminate in a severe disturbance in speech rhythm. Rhythm metrics useful for distinguishing these phrases from those of the other dysarthrias were all related to high temporal variability. Insofar as a function of rhythm is to facilitate tracking and anticipation of word boundaries, random rhythmic disturbance should have deleterious effects on lexical segmentation. This likely explains why, other than severity, the LBE pattern for hyperkinetic speech yielded the lowest $\chi^2$ value of independence, suggesting that the variables' location and type have weaker independent effects on the distribution than was seen for the other groups.

The post hoc attribution of rhythm metrics as explanations for the LBE perceptual outcomes runs the risk of sounding like a just-so story. However, it is important to keep in mind that three of these pools were of equivalent intelligibility with equivalent LBE pool sizes. The only source of explanation for the significant differences in proportions of LBE types across these groups is at the interface of the rhythmic characteristics of the speech signal and listener's use of those characteristics. Understanding that this interface is a tractable source of the intelligibility deficit – something that can be modeled and predicted – provides insight to the communication disorder and plausible interventions[6]. Moreover, for the purposes of

this review, the results emphasize the importance of considering the language-specific linguistic functions of rhythm in studies of intelligibility.

## Crosslinguistic Utilization of English-Centric Descriptions of Dysarthria

In this final section of the review, we return to the use of English-centric descriptors in dysarthrias across languages with an illustration of 'equal and even stress'. Ataxic dysarthria has been well characterized acoustically, perceptually, and kinematically in a number of languages, including English, German, and Swedish [e.g., 73–76]. These are examples of so-called 'stress-timed' languages because of the presence of high durational contrast between stressed and unstressed syllables (table 2). The disruption in timing associated with ataxic dysarthria results in the relative prolongation of normally unstressed syllables, thereby reducing the expected stress-based durational contrasts [3, 75]. The perceptual con sequence of this contrast reduction has been called 'scanning', in which the stream of syllables unfolds with relatively equal and even stress, no longer offering the syllabic cues characteristic of high-contrast languages.

Would one expect the percept of 'scanning' in low-contrast languages whose natural rhythmic structure might already be considered syllabified? Speech rhythm in Japanese is structured around the mora (vocalic or consonant-vocalic pairs), sequences of which are produced with relatively equal durations of constituent vocalic units. Interestingly, the term 'scanning' is regularly used by neurologists to describe the perceptual rhythmic features of ataxic dysarthria in Japanese speakers [e.g., 77–79]. Ikui et al. [80] embarked on a study of the acoustic timing characteristics of ataxic dysarthria in an effort to identify correlates to the perception of scanning in Japanese. They measured mora durations in connected speech and found, as expected, that healthy control speakers produced highly regular mora durations characteristic of the low-stress contrast language. Speakers with ataxic dysarthria, on the other hand, produced highly variable mora durations. They also found a tendency in ataxia to reduce duration of moras with adjacent (double) vowels, which are normally produced with the duration of two moras. The net effect of ataxic dysarthria in Japanese is to interject durational contrasts where none should exist. This presents an interesting language-specific conundrum: the disorder of timing regulation in ataxia serves to *decrease* durational contrasts in English and *increase* durational contrasts in Japanese. As posited by Ikui et al. [80, p. 92]: 'At this point, a question … is whether it is plausible to continue to use the term "scanning" for expressing the abnormal pattern of ataxic speech in Japanese, if the original notion of "scanning" exclusively referred to abnormal "syllabification".'

The quandary implied by the question is one of needing to retrofit the English-centric vocabulary to accommodate a language-specific phenomenon. A more fruitful alternative, however, may be to tip the quandary on its head. If we return to the insight offered by Cutler [81], a universal account of language processing is one that is built from the crosslinguistic details and phenomena between acoustics and perception. This view obviates the need for

---

[6]While a discussion of individual differences in listening strategies is beyond the scope of this review, the topic is of great relevance. The hierarchical segmentation strategy model [67] predicts that listeners will not resort to prosodic cues until the segmental information is sufficiently degraded. This level of degradation will be listener-specific. The interface of rhythmic characteristics and listener strategies is likely to reveal interactions between that interface and the goodness of the speaker's segmental representations.

the language-universal/language-specific distinction because the vantage point is *language processing* rather than acoustic/perceptual features: despite the similarity in the disruption of the acoustics, it is the way in which the necessary perceptual features are disturbed that will dictate the communication impairment.

Figure 2 depicts a generic framework for conceptualizing communication disorders within a language processing approach. The acoustic signatures of the speech signal are ascertained; they are considered relative to the linguistic functions of speech rhythm for a given language, and this converges on predictions regarding the severity and nature of the communication disorder. While segmental goodness – that is, the quality of phoneme articulation – is not explicitly shown in this model, segmental degradation is an implicit component of the acoustic signature.

A critical aspect of this simple model is a language-universal approach to characterizing the acoustic signature. That is, the metrics should capture spectrotemporal prominences and troughs, and their contributors at the source and filter levels of speech production, irrespective of the language and motor speech disorder analyzed. This is important not only to facilitate crosslinguistic investigations, but for the study of the dysarthrias as movement disorders. The language-specific considerations are then framed relative to these metrics to predict the communication disorder outcome.

We are presently developing a package of automated acoustic metrics that will form the basis for crosslinguistic analysis of connected dysarthric speech. This package will include an automated version of the vocalic-intervocalic rhythm metrics [7, 82] ; analysis of the modulation of the amplitude envelope in frequency bands across the signal (envelope modulation spectrum) [15], and analysis of the long-term average spectrum. It also will include metrics developed for the telecommunications industry, which have not been applied previously to disordered speech. The ITU P.563 [83] standard is designed to measure the quality of speech, specifically assessing parameters related to the shape of atypical vocal tracts and the unnaturalness of the vocal quality. This is accomplished through statistical analyses of the cepstral and linear prediction coefficients, which model the source and filter responsible for production of a given speech signal. The result of this multipronged acoustic analysis is a high level characterization of the speech signal that will have two important applications. First, analyses across languages and dysarthrias would reveal commonalities in manifestation of speech motor impairments that may have diagnostic or localizing value, without reference to the *communication disorders* they produce. This retains intention of the 'movement disorder' perspective of the Mayo Classification System, but without invoking (or relying upon) language-specific perceptual descriptors [see ref. 84, for a discussion of taxonomical phenomena in motor speech disorders as an alternative to perceptual classification]. Second, the acoustic characterization forms the basis for consideration relative to the language-specific cue use, as shown in figure 2.

Thus, the second step in this model is to delineate the language-specific linguistic functions of the spectrotemporal prominences and troughs. Many of these details already are available in the literature, and table 2 offers a starting point for communication disorder outcome predictions with regard to lexical segmentation. Returning to our Japanese example, we can

ask what are linguistic functions of rhythm cues and how are they used by native communicators. Otake et al. [85] demonstrated that Japanese listeners use the rhythm unit of moras – rather than the syllables in which they often coexist – to drive lexical segmentation. Lexical segmentation, then, would be compromised in Japanese to the extent that pitch and duration cues to mora identity (or psychological realization) are degraded or distorted. Therapeutic (or automated) improvements of these cues would be justified as an intervention target to improve lexical segmentation and speech intelligibility [see ref. 86 for a cognitive-perceptual framework for intervention in hypokinetic dysarthria].

## Conclusion

Although the Darley et al. [4] classification of dysarthria types is used by clinicians and scientists around the world, there are compelling reasons to assume the dysarthria types and their effect on communication impairment are not consistent across different languages. This review presents a theoretical framework for how well-attested rhythmic differences across languages are likely to affect the nature and possibly degree of the communication impairment in motor speech disorders. Data collected at Arizona State University, USA, and published over the last 15 years, demonstrate that even within a single language – American English – the nature of LBEs made by listeners varies depending on the details of speech rhythm disorders associated with neurological disease. LBEs are shown to be different among persons classified with hypokinetic, ataxic, and mixed dysarthria, all of who are native speakers of English. These LBEs and their effect on speech intelligibility are attributed to the unique speech rhythm disturbances among the three dysarthria types. If this effect can be demonstrated in speakers of the same language, continued investigation on the effect of dysarthria on lexical boundary detection in particular, and speech intelligibility in general, is absolutely essential for languages with rhythmic structures different from English. Such research can only enhance the theory of dysarthria by moving away from an English-centric view of the disorder to a more universal understanding of the effects of dysarthria on communication. The theoretical and data-based considerations presented in this paper also show why it is critical to appreciate the tight link between speech production characteristics on the one hand, and language perception strategies on the other hand, for a full understanding of how motor speech disorders affect communication.

## Acknowledgments

## Appendix

## Methodological Details of the Study

### Participants

**Speakers**—In the large-scale study of which this experiment was a part, 344 potential participants with neurological disease were screened for inclusion. Of these, 269 were not eligible because they did not present with a dysarthria of at least moderate severity, or the severity of their illnesses prevented them from being able to participate. The remaining 75

provided speech samples, and 44 of these (6 men and 6 women from each dysarthria group, with the exception of the hypokinetic group, in which 2 women and 6 men were used) were selected for inclusion in the study, based on the quality and character of their speech deficits exhibited for the experimental phrases. A group of 5 neurologically healthy participants also provided speech samples for this investigation. Because they did not present with intelligibility deficits, their samples were used only for a point of comparison for acoustic measures.

**Listeners**—Participants were 60 undergraduate and graduate students enrolled in classes at Arizona State University. All were native English speakers between the ages of 18 and 47 (mean age = 26, SD = 7.1) who self-reported normal hearing and no disease or conditions known to affect speech or language processing. The group was predominantly female (n = 52) with 8 males. Listeners were compensated for their participation.

## Speech Material

Eighty-six-syllable phrases were developed to elicit and evaluate lexical segmentation errors. These phrases were similar in structure to those of Liss et al. [68] and were comprised of 3–5 words, with alternating syllabic strength [after ref. 65]. Strong syllables (S) were those that would be produced in citation form with a full (unreduced) vowel, and receive primary lexical stress. Weak syllables (W) were those that would be produced with a reduced vowel (schwa or schwar) and without lexical stress [94]. Half of the phrases were of a trochaic stress pattern (SW), and the other half iambic (WS).

All phrases contained either mono- or disyllabic American English words characterized as having low interword predictability (except for articles), such that contextual cues were not of use for word identification. Individual words were selected to be in the moderate word frequency range, as defined by Kucera and Francis [95]. However, a small set of monosyllabic words that served as weak syllables (mostly articles and pronouns) were of high word frequency and they appeared no more than 5 times each across the phrases. In addition, each phrase contained 0–2 target syllables that were designed to, cumulatively, assess a variety of phonemic contrasts [based on ref. 96]. The data from the segmental analysis will be presented in a separate paper.

## Speech Sample Collection, Editing, and Acoustic Measures

Participants were fitted with a head-mount microphone (Plantronics DSP-100), seated in a sound-attenuating booth, and they read stimuli from visual prompts on a computer screen. Recordings were made using a custom script in TF32 [97] (16-bit, 44 kHz) and were saved directly to disk for subsequent editing using commercially available software (SoundForge, Sony, 2004). Several sets of speech material were collected during the session, including the 80 phrases, which serve as the basis for this report. Participants were encouraged to speak in their normal, conversational voice.

Following the collection of all phrases for a speaker group, 8–12 phrases from each participant were selected to create the 80-phrase experimental set. Phrases were selected based on the presence of distinguishing features consistent with the dysarthria subtype (table

1). In all cases, two certified speech-language pathologists identified a subset of suitable phrases for each speaker independently. Common selections were considered options for inclusion in the experimental set. Ultimately, 4 listening sequences were constructed, one for each dysarthria group, each consisting of all the 80 phrases.

Standard acoustic measures of total duration, vowel duration, first and second formant frequencies of strong vowels, and standard deviation of fundamental frequency across each utterance were made using TF32. The phrases were also segmented via Pratt script into vocalic and intervocalic units, which formed the basis of a speech rhythm analysis [see ref. 6].

### Study Protocol and Data Collection

Transcription data were collected using Alvin, a stimulus-presentation software package designed for speech perception experiments [98]. Phrases were randomly presented in an open transcription task with no feedback or replay capabilities.

All listeners performed the task wearing sound-attenuating Sennheiser HD 25 SP headphones in a quiet room in listening carrels, which minimized visual distractions. At the beginning of the experiment, the signal volume was set to a comfortable listening level by each listener, and remained at that level for the duration of the task. After instructions to write down exactly what they heard, listeners were presented with 3 practice phrases, produced by a healthy control speaker. These practice phrases did not contain any of the same words found in the 80 experimental phrases. Performance confirmed understanding of the task for all listeners.

Following the practice phrases, listeners heard each experimental phrase and immediately transcribed what they had heard. The listeners were told they would hear phrases produced by people with diseases that affected the clarity of their speech, but that all the words in the phrases were real English words. They were encouraged to guess on a word if they were not entirely sure about what they heard, or to use an X or dash for parts of the phrase that were not understood.

### Transcript Coding and Reliability

**Lexical Segmentation—**The transcripts were coded independently by two trained judges for the number of words correctly transcribed as well as the presence, type (insertion or deletion), and location (before strong or before weak syllables) of LBEs. In addition, transcription errors that did not violate lexical boundaries were tallied (word substitutions) as well as instances in which no attempt was made at transcription of a phrase (no response). It should be noted that word substitutions and LBEs are mutually exclusive. Word substitution errors occur when the transcription is correctly parsed (thus no LBE) but the response is not the target word. For example, a listener's response of 'advice' for the target 'convince' is a word substitution, whereas the response 'the fence' is an insertion of boundary before a strong syllable. Coding discrepancies were addressed with revision or exclusion from the analysis. Criteria for scoring words correct were identical to those in Liss et al. [68], and included tolerance of word-final morphemic alterations, which did not affect

the number of syllables (i.e., 'boats' for 'boat', but not 'judges' for 'judge'), as well as substitutions of function words 'a' and 'the'. Examples of listener transcripts and coding of LBEs are shown in table 2.

### Data Analysis

The dependent variables, tallied for each listener and group, were: (1) the number of words correctly transcribed out of the total number of words possible (intelligibility score), (2) total number of word transcription errors that did not violate lexical boundaries (word substitutions), and (3) the number, type (insertion vs. deletion), and location (before strong vs. before weak syllables) of LBEs. Accordingly, LBEs fell in four categories: insertion of a word boundary before a strong syllable (IS), insertion of a word boundary before a weak syllable (IW), deletion of a word boundary before a strong syllable (DS), deletion of a word boundary before a weak syllable (DW). For the purpose of comparing patterns across the dysarthria types, an MSS ratio, defined as the number of MSSconsistent LBEs, namely, insertions of word boundaries before strong syllables and deletions of word boundaries before weak syllables, divided by the total number of LBEs (IS + DW/total LBEs). An MSS ratio greater than 0.50 was taken as evidence of stress-based segmentation, because the opportunities to commit nonpredicted errors were greater than predicted errors.

Individual one-way analyses of variances were conducted across dysarthria groups to compare the equality of means of the following dependent variables: percent intelligibility, and number of LBEs. Pairwise comparisons with a Bonferroni correction were also conducted.

$\chi^2$ tests of independence were conducted within groups to test the null hypothesis that the variables of LBE type (insertion or deletion) and location (before strong and before weak) were independent. Significant findings were interpreted as meaning that the variables were dependent, in support of the MSS hypothesis.

$\chi^2$ goodness of fit tests were conducted to compare the LBE distributions between groups to test the null hypothesis that they were drawn from the same sample. Significant findings were interpreted as indicating that the distributions were different for the two dysarthria groups.

## References

1. Kohler K. Rhythm in speech and language: a new research paradigm. Phonetica. 2009; 66:29–45. [PubMed: 19390229]

2. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. J Speech Hear Res. 1969; 12:246–269. [PubMed: 5808852]

3. Darley FL, Aronson AE, Brown JR. Clusters of deviant speech dimensions in the dysarthrias. J Speech Hear Res. 1969; 12:462–496. [PubMed: 5811846]

4. Darley, FL.; Aronson, AE.; Brown, JR. Motor Speech Disorders. Boston: Little Brown; 1975.

5. Duffy, JR. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. St Louis: Mosby-Year Book; 1995.

6. Liss JM, White L, Mattys SL, Lansford K, Lotto AJ, Spitzer SM, Caviness JN. Quantifying speech rhythm abnormalities in the dysarthrias. J Speech Lang Hear Res. 2009; 52:1334–1352. [PubMed: 19717656]

7. Ma J, Whitehill T, So S. Intonation contrast in Cantonese speakers with hypokinetic dysarthria associated with Parkinson's disease. Speech Lang Hear Res. 2010; 53:836–849.

8. Whitehill TL. Studies of Chinese speakers with dysarthria: informing theoretical models. Folia Phoniatr Logop. 2010; 62:92–96. [PubMed: 20424463]

9. Letter MD, Santens P, Estercam I, Van Maele G, De Bodt M, Boon P, et al. Levodopa-induced modifications of prosody and comprehensibility in advanced Parkinson's disease as perceived by professional listeners. Clin Linguist Phon. 2007; 21:783–791. [PubMed: 17882694]

10. Mori H, Kobayashi Y, Kasuya H, Hirose H. Evaluation of fundamental frequency (F0) characteristics of speech in dysarthrias: a comparative study. Acoust Sci Technol. 2005; 26:540–543.

11. Yuceturk A, Yilmaz H, Egrilmez M, Karaca S. Voice analysis and videolaryngostroboscopy in patients with Parkinson's disease. Laryngology. 2002; 259:290–293.

12. Chakraborty N, Roy T, Hazra A, Biswas A, Bhattacharya K. Dysarthric Bengali speech: a neurolinguistic study. J Postgrad Med. 2008; 54:268–272. [PubMed: 18953144]

13. Kotz SA, Schwartze M. Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends Cogn Sci. 2010; 14:392–399. [PubMed: 20655802]

14. Pike, KL. The Intonation of American English. Ann Arbor: University of Michigan Press; 1945.

15. Liss JM, Legendre S, Lotto AJ. Discriminating dysarthria type from envelope modulation spectra. J Speech Lang Hear Res. 2010; 53:1246–1255. [PubMed: 20643800]

16. Loukina A, Kochanski G, Rosner B, Shih C, Keane E. Rhythm measures and dimensions of durational variation in speech. J Acoust Soc Am. 2011; 129:3258–3270. [PubMed: 21568427]

17. Wiget L, White L, Schuppler B, Grenon I, Rauch O, Mattys SL. How stable are acoustic metrics of contrastive speech rhythm? J Acoust Soc Am. 2010; 127:1559–1569. [PubMed: 20329856]

18. Arvaniti A. The usefulness of metrics in the quantification of speech rhythm. J Phonet. 2012; 40:351–373.

19. Dellwo, V. Influences of Speech Rate on the Acoustic Correlates of Speech Rhythm: An Experimental Phonetic Study Based on Acoustic and Perceptual Evidence. Bonn: PhD diss University of Bonn; 2010. Electronic publication: http://hss.ulb.uni-bonn.de:90/2010/2003/2003.htm).

20. White L, Mattys SL, Wiget L. Language categorization by adults is based on sensitivity to durational cues, not rhythm class. J Mem Lang. 2012; 66:665–679.

21. Dauer RM. Stress-timing and syllable-timing reanalyzed. J Phonet. 1983; 11:51–62.

22. Grabe, E.; Low, EL. Durational variability in speech and the rhythm class hypothesis. In: Warner, N.; Gussenhoven, C., editors. Papers in Laboratory Phonology. Vol. 7. Berlin: Mouton de Gruyter; 2002. p. 515-546.

23. Ramus F, Nespor M, Mehler J. Correlates of linguistic rhythm in the speech signal. Cognition. 1999; 73:265–292. [PubMed: 10585517]

24. Arvaniti A. Rhythm, timing and the timing of rhythm. Phonetica. 2009; 66:46–63. [PubMed: 19390230]

25. Cumming R. The language-specific interdependence of tonal and durational cues in perceived rhythmicality. Phonetica. 2011; 68:1–25. [PubMed: 21804331]

26. Niebuhr O. F0-based rhythm effects on the perception of local syllable prominence. Phonetica. 2009; 66:95–112. [PubMed: 19390233]

27. Nolan F, Asu EL. The Pairwise Variability Index and coexisting rhythms in language. Phonetica. 2009; 66:64–77. [PubMed: 19390231]

28. Cutler A, Dahan D, Van Donselaar W. Prosody in the comprehension of spoken language: a literature review. Lang Speech. 1997; 40:141–202. [PubMed: 9509577]

29. Lehiste, I. Suprasegmentals. Cambridge: MIT Press; 1970.

30. Lehiste, I. The phonetic structure of paragraphs. In: Nooteboom, S.; Cohen, A., editors. Structure and Process in Speech Perception. Berlin: Springer; 1980. p. 195-206.

31. Langus A, Marchetto E, Bion RAH, Nespor M. The role of prosody in discovering hierarchical structure in continuous speech. J Mem Lang. 2012; 66:285–306.

32. Rothermich K, Schmidt-Kassow M, Kotz SA. Rhythm's gonna get you: regular meter facilitates semantic sentence processing. Neuropsychologia. 2012; 50:232–244. [PubMed: 22178743]

33. Cummins F. Rhythm as entrainment: the case of synchronous speech. J Phonet. 2009; 37:16–28.

34. Deng S, Srinivasan R. Semantic and acoustic analysis of speech by functional networks with distinct time scales. Brain Res. 2011; 1346:132–144. [PubMed: 20580635]

35. Geiser E, Ziegler E, Jancke L, Meyer M. Electrophysiological correlates of meter and rhythm processing in music perception. Cortex. 2009; 45:93–102. [PubMed: 19100973]

36. Ghitza O. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol. 2011; 2:130. [PubMed: 21743809]

37. Rothermich K, Schmidt-Kassow M, Schwarze M, Kotz SA. Event-related potential responses to metric violations: rules versus meaning. Neuroreport. 2010; 21:580–584. [PubMed: 20431495]

38. Snyder JS, Pasinski AC, McAuley JD. Listening strategy for auditory rhythms modulates neural correlates of expectancy and cognitive processing. Psychophysiology. 2011; 48:198–207.

39. Stephens GJ, Silbert LJ, Hasson U. Speakerlistener neural coupling underlies successful communication. Proc Natl Acad Sci USA. 2010; 107:14425–14430. [PubMed: 20660768]

40. Tilsen S. Metrical regularity facilitates speech planning and production. Lab Phonol. 2011; 2:185–218.

41. Zanto TP, Snyder JS, Large EW. Neural correlates of rhythmic expectancy. Adv Cogn Psychol. 2006; 2:221–231.

42. Gill S. Rhythmic synchrony and mediated interaction: towards a framework of rhythm in embodied interaction. AI Soc. 2012; 27:111–127.

43. Kroger BJ, Kopp S, Lowit A. A model for production, perception, and acquisition of actions in face-to-face communication. Cogn Process. 2010; 11:187–205. [PubMed: 20012153]

44. Cutler A. Greater sensitivity to prosodic goodness in non-native than in native listeners. J Acoust Soc Am. 2009; 125:3522–3525. [PubMed: 19507933]

45. Cutler A, Cooke M, Garcia Lecumberri ML, Pasveer D. L2 consonant identification in noise: cross-language comparisons. Proc INTERSPEECH, Antwerp. 2007 Aug.:1585–1588.

46. Kim J, Davis C, Cutler A. Perceptual tests of rhythmic similarity. II. Syllable rhythm. Lang Speech. 2008; 51:343–359. [PubMed: 19348155]

47. Tyler MD, Cutler A. Cross-language differences in cue use for speech segmentation. J Acoust Soc Am. 2009; 126:367–376. [PubMed: 19603893]

48. Weber A, Cutler A. First-language phonotactics in second-language listening. J Acoust Soc Am. 2006; 119:597–607. [PubMed: 16454313]

49. Huang T, Johnson K. Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. Phonetica. 2011; 67:243–267. [PubMed: 21525779]

50. Kinsbourne M, Jordon JS. Embodied anticipation: a neurodevelopmental interpretation. Discourse Processes. 2009; 46:103–126.

51. Kooijman V, Hagoort P, Cutler A. Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. Infancy. 2009; 6:591–612.

52. Liu C, Rodriguez A. Categorical perception of intonation contrasts: effects of listeners' language background. J Acoust Soc Am. 2012; 131:EL427–EL433. [PubMed: 22713017]

53. Hay JF, Pelucchi B, Estes KG, Saffran JR. Linking sounds to meanings: infant statistical learning in a natural language. Cogn Psychol. 2011; 63:93–106. [PubMed: 21762650]

54. Franco A, Cleeremans A, Destrebecqz A. Statistical learning of two artificial languages presented successively: how conscious? Front Psychol. 2011; 2:1–12. [PubMed: 21713130]

55. McQueen, JM.; Cutler, A. Cognitive processes in speech perception. In: Hardcastle, WJ.; Laver, J.; Gibbon, FE., editors. The Handbook of Phonetic Sciences. ed 2. Oxford: Blackwell; 2010. p. 489-520.

56. Oller DK. The effect of position in utterance on speech segment duration in English. J Acoust Soc Am. 1973; 54:1235–1247. [PubMed: 4765808]

57. Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price P. Segmental durations in the vicinity of prosodic phrase boundaries. J Acoust Soc Am. 1992; 91:1707–1717. [PubMed: 1564206]

58. Klatt DH. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J Acoust Soc Am. 1976; 59:1208–1220. [PubMed: 956516]

59. Turk AE, White L. Structural influences on accentual lengthening in English. J Phonet. 1999; 27:171–206.

60. Fry DB. Duration and intensity as physical correlates of linguistic stress. J Acoust Soc Am. 1955; 27:765–768.

61. Lieberman P. Some acoustic correlates of word stress in American English. J Acoust Soc Am. 1960; 32:451–454.

62. Fear B, Cutler A, Butterfield S. The strong/weak syllable distinction in English. J Acoust Soc Am. 1995; 9:1893–1904. [PubMed: 7699171]

63. Spitzer S, Liss JM, Mattys SL. Acoustic cues to lexical segmentation: a study of resynthesized speech. J Acoust Soc Am. 2007; 122:3678–3687. [PubMed: 18247775]

64. Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. J Exp Psychol Hum Percept Perform. 1988; 14:113–121.

65. Cutler A, Butterfield S. Rhythmic cues to speech segmentation: evidence from juncture misperception. J Mem Lang. 1992; 31:218–236.

66. Smith MR, Cutler A, Butterfield S, Nimmo-Smith I. The perception of rhythm and word boundaries in noise-masked speech. J Speech Hear Res. 1989; 32:912–920. [PubMed: 2601320]

67. Mattys SL, White L, Melhorn JF. Integration of multiple speech segmentation cues: a hierarchical framework. J Exp Psychol Gen. 2005; 134:477–500. [PubMed: 16316287]

68. Liss JM, Spitzer S, Caviness JN, Adler C, Edwards B. Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. J Acoust Soc Am. 1998; 104:2457–2466. [PubMed: 10491707]

69. Liss JM, Spitzer SM, Caviness JN, Adler C. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. J Acoust Soc Am. 2002; 112:3022–3030. [PubMed: 12509024]

70. Mattys SL, Davis MH, Bradlow AR, Scott SK. Speech recognition in adverse conditions: a review. Lang Cogn Processes. 2012; 27:953–978.

71. Choe Y, Liss JM, Azuma T, Mathy P. Evidence of cue use and performance differences in deciphering dysarthric speech. J Acoust Soc Am. 2012; 131:EL112–EL118. [PubMed: 22352609]

72. Liss JM, Spitzer SM, Caviness JN, Adler C, Edwards BW. Lexical boundary error analysis in hypokinetic and ataxic dysarthria. J Acoust Soc Am. 2000; 107:3415–3424. [PubMed: 10875386]

73. Ackerman H, Hertrich I. Speech rate and rhythm in cerebellar dysarthria: an acoustic analysis of syllabic timing. Folia Phoniatr Logop. 1994; 46:70–78. [PubMed: 8173615]

74. Hartelius L, Runmarker B, Andersen O, Nord L. Temporal speech characteristics of individuals with multiple sclerosis and ataxic dysarthria: 'scanning speech' revisited. Folia Phoniatr Logop. 2000; 52:228–238. [PubMed: 10965176]

75. Kent RD, Netsell R, Abbs JH. Acoustic characteristics of dysarthria associated with cerebellar disease. J Speech Hear Res. 1979; 22:627–648. [PubMed: 502519]

76. Ziegler W, Wessel K. Speech timing in ataxic disorders: sentence production and rapid repetitive articulation. Neurology. 1996; 47:208–214. [PubMed: 8710079]

77. Okawa S, Sugawara M, Watanabe S, Imota T, Toyoshima I. A novel sacsin mutation in a Japanese woman showing clinical uniformity of autosomal recessive spastic ataxia of Charlevoix-Saguenay. J Neurol Neurosurg Psychiatry. 2006; 77:280–282. [PubMed: 16421146]

78. Takahashi Y, Miyajima H, Shirabe S, Nagataki S, Suenaga A, Gitlin JD. Characterization of a nonsense mutation in the ceruloplasmin gene resulting in diabetes and neurodegenerative disease. Hum Mol Genet. 1996; 5:81–84. [PubMed: 8789443]

79. Takahashi T, Igarashi S, Kimura T, Hozumi I, Kawachi I, Onodera O, Takano H, Saito M, Tsuji S. Japanese cases of familial hemiplegic migraine with cerebellar ataxia carrying a T666M mutation in the CACNA1A gene. J Neurol Neurosurg Psychiatry. 2002; 72:676–677. [PubMed: 11971066]

80. Ikui Y, Tsukuda M, Kuroiwa Y, Koyano S, Hirose H, Taguchi T. Acoustic characteristics of ataxic speech in Japanese patients with spinocerebellar degeneration (SCD). Int J Lang Commun Disord. 2012; 47:84–94. [PubMed: 22268904]

81. Cutler A. The comparative perspective on spoken-language processing. Speech Commun. 1997; 21:3–15.

82. Utianski RL, Berisha V, Liss JM. The sensitivity of P.563 features to idiosyncratic patterns of speech degradation. J Speech Lang Hear Res. under review.

83. Malfait L, Berger J, Kastner M. P.563 – the ITU-T standard for single-ended speech quality assessment. IEEE Trans Audio Speech Lang Process. 2006; 14:1924–1934.

84. Weismer, G.; Kim, Y-J. Classification and taxonomy of motor speech disorders: what are the issues?. In: Maassen, B.; van Lieshout, PHHM., editors. Speech Motor Control: New Developments in Basic and Applied Research. Cambridge: Oxford University Press; 2010. p. 229-241.

85. Otake T, Hatano G, Cutler A, Mehler J. Mora or syllable? Speech segmentation in Japanese. J Mem Lang. 1993; 32:258–278.

86. Lansford KL, Liss JM, Caviness JN, Utianski RL. A cognitive-perceptual approach to conceptualizing speech intelligibility deficits and remediation practice in hypokinetic dysarthria. Parkinsons Dis. 2011

87. Cutler, A.; Pasveer, D. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In: Hoffman, R.; Mixdorff, H., editors. 3rd Int Conf on Speech Prosody. Dresden: TUDpress; 2006. p. 237-240.

88. Suomi K, McQueen JM, Cutler A. Vowel harmony and speech segmentation in Finnish. J Mem Lang. 1997; 36:422–444.

89. Fant G, Kruckenberg A. Notes on stress and word accent in Swedish. Proc Int Symp Prosody. 1994:2–3.

90. Spinelli E, McQueen JM, Cutler A. Processing resyllabified words in French. J Mem Lang. 2003; 48:233–254.

91. Perry C, Wong RK-S, Matthews S. Syllable timing and pausing: evidence from Cantonese. Lang Speech. 2009; 52:29–53. [PubMed: 19334415]

92. Frota S, Vigàrio M. On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. Probus. 2001; 13:247–275.

93. Spitzer SM, Liss JM, Caviness JN, Adler C. An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech. J Med Speech Lang Pathol. 2000; 8:285–294.

94. Van Oijen B. Vowel mutability and lexical selection in English: evidence from a word reconstruction task. Mem Cognit. 1996; 24:573–583.

95. Kucera, H.; Francis, WN. Computational Analysis of Present-Day American English. Providence: Brown University Press; 1967.

96. Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. J Speech Hear Disord. 1989; 54:482–489. [PubMed: 2811329]

97. Milenkovic P. TF32 (computer software). Madison, University of Wisconsin-Madison, Department of Electrical and Computer Engineering. 2004

98. Hillenbrand JM, Gayvert RT. Open source software for experiment design and control. J Speech Lang Hear Res. 2005; 48:45–60. [PubMed: 15938059]
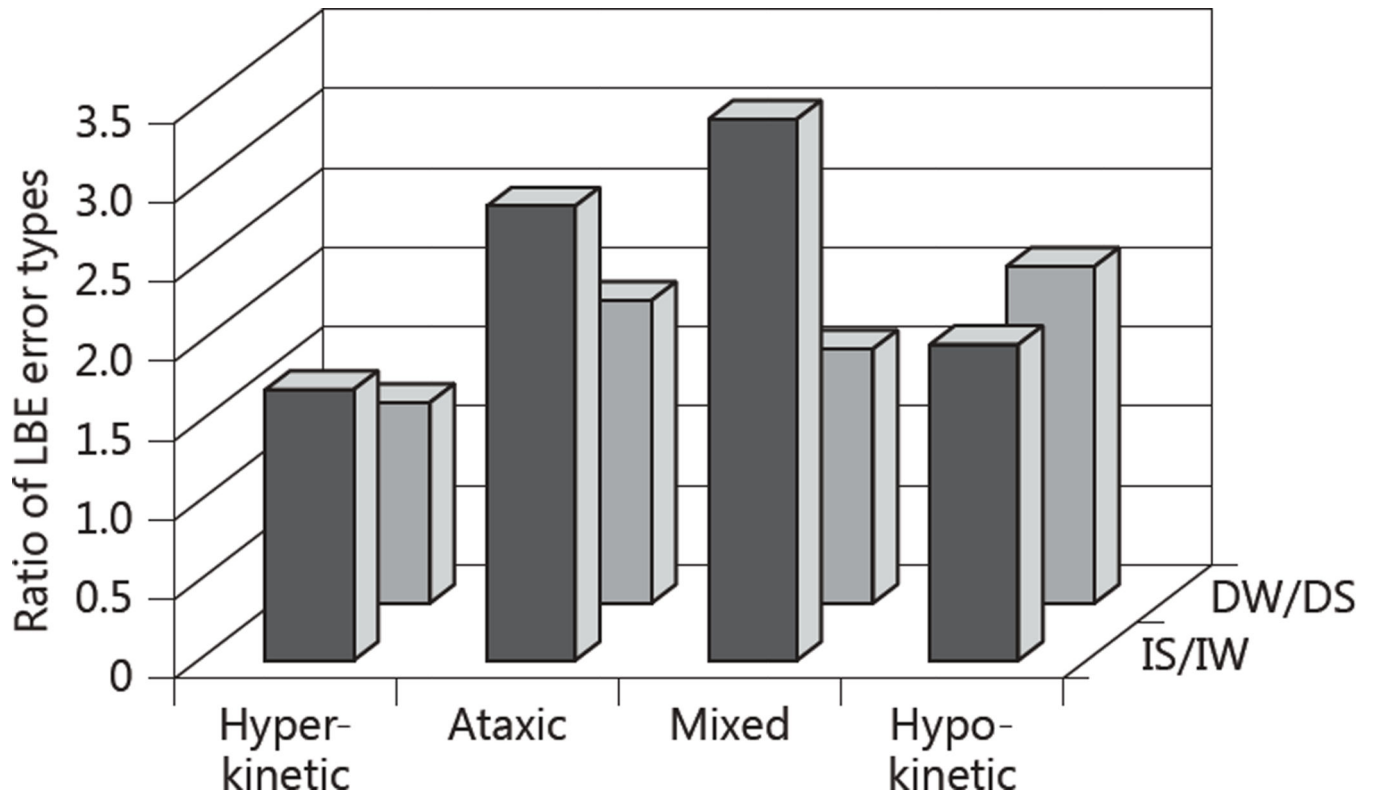
**Figure 1.**
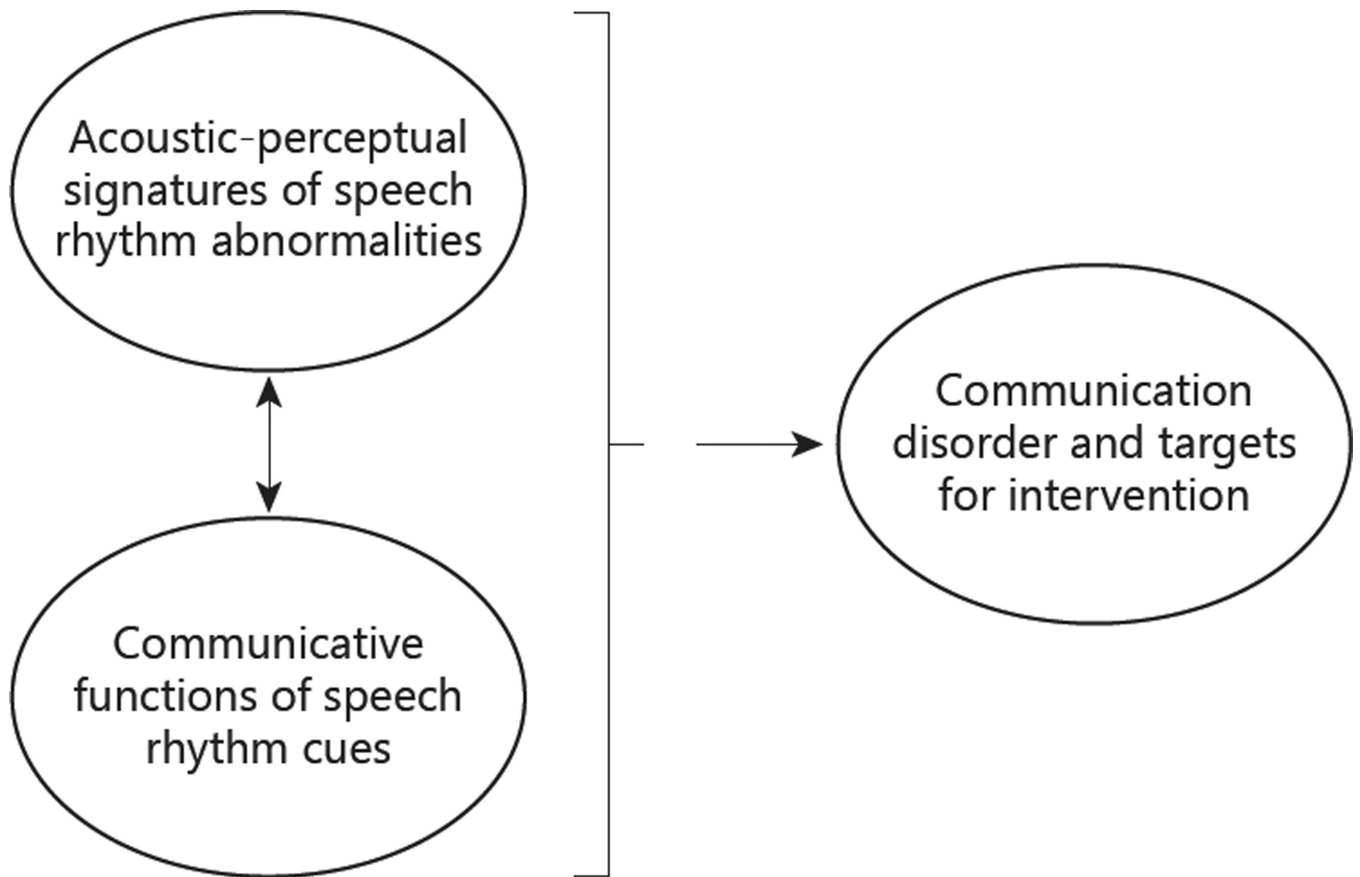Ratio of predicted to nonpredicted LBE types for each of the four dysarthria subtypes.

**Figure 2.**
Schematic of generic framework for a language processing approach.

**Table 1**

List of languages and dialects applying English-centric rhythm descriptors in dysarthrias

| |
| --- |
| Australian English (n = 3) |
| British English (n = 7) |
| Belgian (n = 1) |
| Bengali (n = 1) |
| Canadian French (n = 1) |
| Cantonese (n = 3) |
| Dutch (n = 2) |
| Farsi (n = 2) |
| French (n = 3) |
| German (n = 5) |
| Greek (n = 1) |
| Hebrew (n = 1) |
| Hindi (n = 1) |
| Italian (n = 1) |
| Japanese (n = 4) |
| Mandarin (n = 3) |
| New Zealand English (n = 1) |
| Portuguese (Portugal) (n = 1) |
| Portuguese (Brazil) (n = 1) |
| South African English (n = 1) |
| Spanish (n = 5) |
| Swedish (n = 1) |
| Thai (n = 2) |

n = Number of articles in each language identified in a search of the literature. The list of articles can be found in online supplementary 'Appendix A'.

**Table 2**

Cues of rhythm used for segmenting word and syllable boundaries

| Language | Timing class | Cues for segmentation |
|----------|--------------|------------------------|
| English | stress | duration<br>loudness<br>F0 variation<br>vowel quality |
| Dutch | stress | suprasegmental cues to stress<br>vowel quality (less salient distinction than in English) [87]<br>pitch movement cues [47] |
| Finnish | stress | vowel harmony [88]<br>durational contrast of stressed vowel |
| German | stress | final-syllable lengthening (regardless of stress) [1] |
| Swedish | stress | word accent fall (F0, duration, loudness) [89]<br>sentence accent rise |
| French | syllable | durational contrast of consonants [90] |
| Spanish | syllable | fine-tuned discrimination to final lengthening [20]<br>stress placement |
| Cantonese | syllable | different silent pause intervals [91] |
| Mandarin | mixed [syllable and mora (tone)] | lexical tone, determined by F0 height and F0 contour<br>tone duration, relative to sentence position |
| Brazilian Portuguese | mixed (mora and syllable) [92] | reduced vowels in unstressed position<br>simplified consonant clusters<br>rate-dependent changes in durational contrasts |
| Japanese | mora (tone) | subsyllabic segmentation [92] |

F0 = Fundamental frequency.

**Table 3**

Summary of published findings related to lexical segmentation of dysarthric speech from the Motor Speech Disorders Laboratory at Arizona State University

| Speaker population | Analysis type(s) | Findings | Publication |
|---|---|---|---|
| Hypokinetic | LBE analysis<br>Acoustic measures of syllable strength | All listeners used syllabic strength for lexical segmentation decisions<br>Strategy was less effective for severely impaired speech and reduced strength cues | Liss et al. [68] |
| Hypokinetic<br>Ataxic | LBE analysis<br>Acoustic measures of syllable strength | Replicated hypokinetic findings from our 1998 study [68]<br>Ataxic dysarthria did not elicit predicted LBE patterns<br>Concluded prosodic disturbance in ataxic dysarthria renders metrical segmentation difficult | Liss et al. [72] |
| Hypokinetic<br>Ataxic | LBE analysis | Replicated previous hypokinetic and ataxic findings from Liss et al. [68, 72]<br>Brief familiarization benefited all listeners<br>Greater dysarthria-specific than dysarthria-general benefits, but overall no change in LBE patterns<br>Concluded that 'learning' associated with familiarization may be at the segmental level | Liss et al. [72] |
| Hypokinetic<br>Ataxic | Word substitution analysis | In a reanalysis of Liss et al.'s [72] data, word substitutions were segmentally closer to the targets in transcriptions of ataxic as compared with hypokinetic dysarthric speech<br>This offers evidence that familiarization may be learning at the segmental level | Spitzer et al. [93] |
| Resynthesized speech | LBE analysis | Healthy control speech resynthesized to approximate dysarthria-like prosodic patterns<br>Conditions of flattened F0 and of reduced second formant toward a schwa in full vowels resulted in the greatest impediment to implementing a metrical segmentation strategy<br>Findings did not align completely with expectations for duration-cue reductions and ataxic speech predictions | Spitzer et al. [63] |
| Healthy control<br>Hypokinetic<br>Ataxic<br>Hyperkinetic<br>Mixed spastic-flaccid | Temporally based rhythm metrics<br>DFA | DFAs distinguished rhythm metrics for healthy control speech from those of dysarthric speech<br>Rhythm metrics reliably classified dysarthrias into their categories with good accuracy | Liss et al. [6] |
| Healthy control<br>Hypokinetic<br>Ataxic<br>Hyperkinetic<br>Mixed spastic-flaccid | EMS<br>DFA | EMS, an automated spectral analysis of the low-rate amplitude modulations of the envelope for the entire speech signal and within select frequency bands, performed as well as the hand-measured vocalic and intervocalic interval durations employed in Liss et al. [6] | Liss et al. [15] |
| Mixed spastic-flaccid | LBE analysis | Data show evidence for cue use differences among better and poorer listeners | Choe et al. [71] |

DFA = Discriminant function analysis; EMS = envelope modulation spectrum.

**Table 4**

Mean percent of words correct, total number of LBEs, and distribution of LBEs (including percent of total occurrence of LBEs) per dysarthria group

| Dysarthria | Mean percent of words correct (SE) | Total LBEs n | IS | IW | DS | DW |
|---|---|---|---|---|---|---|
| Hyperkinetic | 36 (0.011) | 957 | *432 (45%)* | 253 (26%) | 120 (13%) | *152 (16%)* |
| Ataxic | 44 (0.015) | 788 | *384 (49%)* | 134 (17%) | 93 (12%) | *177 (22%)* |
| Mixed | 49 (0.01) | 721 | *433 (60%)* | 127 (17.5%) | 62 (8.5%) | *99 (14%)* |
| Hypokinetic | 46 (0.014) | 744 | *339 (45.5%)* | 170 (23%) | 75 (10%) | *160 (9.5%)* |

SE = Standard error. Predicted errors shown in italics.

**Table 5**

Results of multiple comparison analyses

| Dysarthria | Comparison group | Percent of words correct mean difference | Number of LBEs mean difference |
|---|---|---|---|
| Hyperkinetic | ataxic<br>mixed<br>hypokinetic | −0.076040[*]<br>−0.124573[*]<br>−0.098040[*] | 11.2667<br>15.7333[*]<br>14.2000[*] |
| Ataxic | hyperkinetic<br>mixed<br>hypokinetic | 0.076040[*]<br>−0.048533<br>−0.022000 | −11.2667<br>4.4667<br>2.9333 |
| Mixed | hyperkinetic<br>ataxic<br>hypokinetic | 0.124573[*]<br>0.048533<br>0.026533 | −15.7333[*]<br>−4.4667<br>−1.5333 |
| Hypokinetic | hyperkinetic<br>ataxic<br>mixed | 0.098040[*]<br>0.022000<br>−0.026533 | 14.2000[*]<br>2.9333<br>1.5333 |

[*] $p < 0.0125$.

**Table 6**

On the diagonal, $\chi^2$ tests of independence (d.f. = 1), and on the off-diagonal, $\chi^2$ goodness of fit tests (d.f. = 3) for LBE types for each of the four dysarthria subtypes

|  | **Hyperkinetic** | **Ataxic** | **Mixed** | **Hypokinetic** |
|---|---|---|---|---|
| Hyperkinetic | 28.64 | 123.05 | 207.64 | 93.44 |
| Ataxic |  | 117.01 | 82.89 | 19.72 |
| Mixed |  |  | 87.53 | 75.27 |
| Hypokinetic |  |  |  | 78.37 |

All results were significant at p < 0.001.

**Table 7**

Correspondence of LBE (perceptual) and rhythm metric (acoustic-durational) findings for each of the four dysarthria subtypes

| Dysarthria | LBE patterns in current report | Corresponding rhythm metric findings [6] |
|---|---|---|
| Hypokinetic | Highest proportion of deletion errors<br>Reduced, but relatively intact, predicted-to-nonpredicted ratios | Rapid speaking rate distinguished hypokinetic speech from all other dysarthric groups<br>Despite reduced intelligibility, temporal relationships among vocalic and consonantal segments – in particular, VarcoVC – showed relative preservation of normal rhythm, as evidenced by the similarity of these scores to those of the control group |
| Ataxic | Relatively preserved predicted-to-nonpredicted ratios | Metrics of variability (VarcoC, rPVI-VC and nPVI-V) measures accurately classified 85% of the ataxic speakers; however, classification errors were primarily with controls as their values were similar |
| Mixed | Highest proportion of insertion errors and lowest deletion errors<br>Highest proportion of insertions before strong syllables | They presented with the slowest speaking rate with greatly prolonged vowels (%V), and the lack of temporal distinction between vowels produced in stressed versus unstressed syllables (nPVI-V, VarcoV)<br>Despite lack of temporal distinction between strong and weak vowels, intonation contours were relatively preserved [variation in F0, not reported in ref. 6] |
| Hyperkinetic | Lowest adherence to metrical segmentation strategy, with reduced predicted-to-nonpredicted ratios | Metrics sensitive to the high variability in the consonantal intervals (VarcoC and VarcoVC) were important for distinguishing hyperkinetic speech from the other dysarthrias; other variables that also captured this variability included C, rPVI-C, rPVI-VC, nPVI-C |