Perception of Coarticulatory Information in Normal Speech and Dysarthria

Kris Tjaden
Joan Sussman
University at Buffalo, Buffalo, NY

Purpose: This study addressed three research questions: (a) Can listeners use anticipatory vowel information in prevocalic consonants produced by talkers with dysarthria to identify the upcoming vowel? (b) Are listeners sensitive to interspeaker variation in anticipatory coarticulation during prevocalic consonants produced by healthy talkers and/or talkers with dysarthria, as measured by vowel identification accuracy? (c) Is interspeaker variation in anticipatory coarticulation reflected in measures of intelligibility?
Method: Stimuli included 106 CVC words produced by 20 speakers with either Parkinson's disease or multiple sclerosis or by 16 healthy controls characterized by an operationally defined normal, under, or over level of anticipatory vowel coarticulation. Ten listeners were presented with prevocalic consonants for identification of the vowel. Ten additional listeners judged single-word intelligibility. An analysis of variance was used to determine differences in vowel identification accuracy and intelligibility as a function of speaker group, coarticulation level, and vowel type.
Results: Listeners accurately identified vowels produced by all speaker groups from the aperiodic portion of prevocalic consonants, but interspeaker variations in strength of coarticulation did not strongly affect vowel identification accuracy or intelligibility.
Conclusions: Listeners appear to be tuned to similar types of information in the acoustic speech stream irrespective of the source or speaker, and any perceptual effects of interspeaker variation in coarticulation are subtle.

KEY WORDS: coarticulation, perception, dysarthria

The perceptual system is faced with the extraordinary task of making sense of a coarticulated speech stream, and theories of speech perception and spoken word recognition have addressed the problem of coarticulation in two primary ways (see reviews in Goldinger, Pisoni, & Luce, 1996; McQueen & Cutler, 1997). One class of theories, including the motor theory of perception (Liberman & Mattingly, 1989), acknowledged the difficulties of listeners in extracting individual speech sounds due to the coarticulated or "encoded" nature of speech. Liberman and Mattingly suggested that as many segments as possible were downloaded to the production system in parallel, resulting in an enmeshed acoustic stream. They believed that listeners resorted to retrieval of invariant articulatory gestures instead of retrieving the intended sounds from the acoustic signal. Empirical support for this class of theories is lacking, however, in part because the search for invariant aspects of the speech signal has not been successful (but see Sussman, Fruchter, Hilbert & Sirosh, 1998).

Another class of theories considers coarticulatory cues as potentially valuable sources of information for perception. TRACE (Elman &

1092-4388/06/4904-0888

McClelland, 1986; McClelland & Elman, 1986) and the distributed cohort theory (e.g., Gaskell & Marslen-Wilson, 1997, 1999) are examples of spoken word recognition theories. TRACE attempts to represent coarticulation by allowing phonetic features to overlap in their representation, even though the features themselves never change. While most word recognition models still do not incorporate most of the contextual and temporal detail of speech signals into their representations (Luce & McClellan, 2005), allophonic variation is incorporated into other recognition models (e.g., PARSYN; Luce, Goldinger, Auer, & Vitevitch, 2000).

As suggested in the following paragraphs, two areas of study show the importance of coarticulation. One set of studies indicates that listeners are sensitive to coarticulatory information in the acoustic speech stream. Other studies show that deviant or inappropriate coarticulatory information is detrimental to accurate perception. Because the current study focuses on anticipatory or right-to-left vowel coarticulation within syllables or words, the following review is largely restricted to studies investigating perception of anticipatory vowel cues within words. Coarticulatory cues between syllables or words may have very different perceptual implications. For example, reduced anticipatory coarticulation in a CV syllable can negatively impact perceptual accuracy for the vowel (e.g., Southwood, Dagenais, Sutphin, & Garcia, 1997; Yeni-Komshian & Soli, 1981), but reduced coarticulation between syllables or words is a useful cue for word boundaries and thus helps to constrain lexical access (see Mattys, 2004).

The perceptual relevance of anticipatory vowel information is suggested by a variety of sources, including studies reporting that listeners can identify vowels in stop consonant + vowel or fricative + vowel syllables produced by healthy adult talkers at better than chance levels, when presented with only the aperiodic portion of the consonant (e.g., Kuehn & Moll, 1972; Nittrouer & Whalen, 1989; Ostreicher & Sharf, 1976; Winitz, Scheib, & Reeds, 1972; Yeni-Komshian & Soli, 1981). Listeners can also identify the vowel in CV syllables produced by adults with apraxia of speech (AOS) or aphasia at better than chance levels when presented with the aperiodic portion of the consonant, at least for those individuals with AOS or aphasia exhibiting unimpaired coarticulatory patterns (Baum, 1998; Katz, 1988). Ostreicher and Sharf (1976) further suggested that because anticipatory cues enable listeners to perceive the identity of later occurring segments, these cues are important for maximizing the efficiency and speed of speech perception.

Studies reporting that experimentally induced coarticulatory deviancies pose difficulty for listeners suggest that coarticulatory cues are a rich information source and thus are important for accurate, efficient speech perception. For example, Martin and Bunnell (1981, 1982) cross-spliced the final syllable /zi/ or /zɑ/ in utterances like /puzi/ and /puzɑ/ and asked listeners to monitor for the second vowel. Listeners were faster and more accurate at identifying a second vowel such as /i/, when the first vowel contained appropriate coarticulatory information for /i/, or as if the first vowel had originally been produced before /i/. On the basis of a series of cross-splicing experiments, Martin and Bunnell (1981, 1982) concluded that appropriate coarticulatory cues facilitate perception while inappropriate anticipatory cues are detrimental to rapid, accurate perception. Relatedly, other studies have cross-spliced stimuli from two languages to create appropriate and inappropriate coarticulatory contexts (Beddor, Harnsberger, & Lindemann, 2002; Manuel, 1987). Beddor et al. (2002) showed that listeners respond to coarticulatory cues in ways consistent with their linguistic experience. The fact that listeners are tuned to the coarticulatory patterns of their native language further indicates the perceptual relevance of coarticulatory cues.

The importance of anticipatory vowel cues for perception is also indicated by studies reporting that naturally occurring coarticulatory deviancies pose difficulty for listeners (Southwood et al., 1987; Waldstein & Baum, 1994; Ziegler & von Cramon, 1986). Southwood et al. (1987), for example, investigated anticipatory vowel coarticulation in CVC words produced by a speaker with AOS and a healthy talker. Speech production measures indicated delayed or distorted anticipation of the vowel during production of the word–initial consonant for the speaker with AOS. A gating paradigm was employed for the perceptual study, wherein listeners were presented with successively lengthened stimuli for identification of the vowel. At earlier gates, vowel identification was better for stimuli produced by the healthy talker, suggesting that the healthy talker's stimuli contained more coarticulatory information for the vowel. Similar results were reported by Ziegler and von Cramon (1985, 1986) for AOS. The perceptually obvious coarticulatory deviancies noted for some talkers with AOS, therefore, might help to explain reduced intelligibility in AOS. The appropriate studies have yet to be conducted, however.

Despite an abundance of studies suggesting the perceptual relevance of coarticulation, much remains to be learned about how and when intrasyllabic coarticulatory information is important for perception. For example, if coarticulatory information is exploited by listeners as posited by some theories of perception, listeners should be tuned to these cues in speech signals for a variety of talkers. Dysarthria provides an interesting case for studying the perceptual relevance of coarticulation. On the one hand, speech production studies suggest similar patterns of anticipatory coarticulation for speakers with dysarthria and healthy controls, at least in studies of mild-to-moderate dysarthria in which between-group differences in articulatory scaling are controlled (Hertrich & Ackermann, 1999; Tjaden, 2003; also Weismer, Yunusova, & Westbury, 2003). Thus, it seems reasonable to speculate that anticipatory vowel cues in the acoustic signal of dysarthria would be perceptible to listeners in the same manner as for neurologically normal speech signals.

On the other hand, articulatory imprecision associated with dysarthria could mask the ability of listeners to effectively use coarticulatory information for perception. For example, Katz, Kripke, and Tallal (1991) found that anticipatory vowel cues in /s/ + vowel syllables produced by 3-year-old children were not perceptible, although speech production measures indicated similar coarticulatory patterns for children and adults (also Sereno, Baum, Katz, & Lieberman, 1987). It was speculated that listeners may have been unable to use anticipatory vowel cues in fricatives extracted from fricative + vowel syllables produced by young children, owing to imprecise consonant production. That is, coarticulatory cues for the vowel were imperceptible due to a poorly produced consonant. Similar statements have been made concerning the influence of poorly articulated vowels on perception of vowel coarticulatory information from prevocalic consonants produced by children with hearing impairment (Waldstein & Baum, 1994). Studies of dysarthria also suggest impaired segmental production of consonants and vowels, at least for some speakers and speech materials (e.g., Tjaden & Turner, 1997; Tjaden & Wilding, 2004). Katz et al. (1991) further noted that young children tend to be less intelligible than older children and adults. One implication is that reduced perceptibility of coarticulatory cues might be related to instances of reduced intelligibility. Indeed, if coarticulatory cues are important for accurate, efficient perception, it stands to reason that there would be consequences if these cues cannot be detected.

Only one study has attempted to investigate perception of coarticulatory cues in the acoustic signal of dysarthria. Ziegler and von Cramon (1985) were primarily interested in studying perception of anticipatory coarticulation for speech produced by an individual with AOS, but listeners also judged stimuli produced by healthy controls and a speaker with spastic dysarthria secondary to a traumatic brain injury. Listeners were able to identify the vowel in CV syllables produced by the healthy talkers when presented with only the consonant, but had difficulty identifying the vowel in stimuli produced by the speaker with dysarthria. Ziegler and von Cramon (1985) suggested that this result was due to the fact that vowels produced by the speaker with dysarthria were poorly distinguished.

Finally, speech production studies have shown that the degree of coarticulation for a given utterance can vary substantially across speakers, even for neurologically normal talkers (e.g., Lubker & Gay, 1982). Theories in which listeners exploit coarticulatory cues for perception offer little in the way of explanation as to whether this type of coarticulatory variability is important. This is likely related to the fact that so few studies have investigated whether listeners are sensitive to this type of "normal" coarticulatory variability. Listeners can tolerate a fair amount of variability in the acoustic signal and still arrive at the same percept (see review of perceptual normalization effects in McQueen & Cutler, 1997). Thus, one possibility is that listeners normalize or compensate for interspeaker variations in coarticulation, as might be required for rate-induced variations in coarticulation (Katz, 2000). As described next, however, at least one study suggests that listeners are sensitive to interspeaker variation in anticipatory vowel coarticulation.

Yeni-Komshian and Soli (1981) presented listeners with the /s/ in /su/ produced by two healthy speakers. Listeners were only able to identify the vowel in stimuli produced by one of the speakers. Acoustic analyses indicated anticipatory lip rounding for /u/ throughout the entire fricative for the speaker for whom vowels were accurately identified, as indicated by a dip in the spectrum in the vicinity of 5 kHz. For the speaker for whom vowels were poorly identified from the aperiodic portion of the consonant, anticipatory lip rounding for /u/, as inferred from a dip in the consonant spectra in the vicinity of 5 kHz, was only evident immediately before vowel onset. These results support the idea that listeners are sensitive to normal amounts of interspeaker variation in intrasyllabic coarticulation. Whether this type of coarticulatory variation is perceptually meaningful, as indexed by a more functional measure such as intelligibility, requires further study. Yeni-Komshian and Soli (1981) noted that syllables produced by both speakers were intelligible when heard in their entirety, but this claim was not substantiated with formal measures of intelligibility.

In summary, if coarticulatory cues are exploited for perception as suggested by certain perceptual theories, listeners should be sensitive to these cues in speech signals for a variety of talkers. It is unknown whether listeners are sensitive to coarticulatory cues in the acoustic signal of dysarthria, although dysarthria provides an interesting case for studying the perceptual relevance of coarticulatory cues. A theory of spoken language perception should also have a full appreciation for all of the varieties of coarticulation that are perceptually relevant. While it is known that listeners are sensitive to interspeaker differences in coarticulation when coarticulatory patterns are deviant for one of two speakers, such as in the Southwood et al. (1987) study comparing listeners' responses to speech signals for a talker with AOS and a neurologically normal talker, the extent to which listeners are sensitive to normal amounts of interspeaker variation in coarticulation is not well established. It also is unknown whether this type of coarticulatory variability is perceptually meaningful, as indexed by a measure such as intelligibility. The current series of experiments sought to further investigate the perceptual relevance of intrasyllabic anticipatory vowel coarticulation by addressing the following research questions:

1. Can listeners use anticipatory vowel information in prevocalic consonants produced by talkers with dysarthria to identify the upcoming vowel?
2. Are listeners sensitive to interspeaker variation in anticipatory vowel coarticulation during prevocalic consonants produced by neurologically normal talkers and/or talkers with dysarthria, as measured by vowel identification accuracy?
3. Is interspeaker variation in anticipatory vowel coarticulation during prevocalic consonants reflected in measures of intelligibility for neurologically normal speakers and/or talkers with dysarthria?

General Method

Stimulus Characteristics

Speakers. Speakers who produced stimuli selected for study included 10 individuals with a neurological diagnosis of multiple sclerosis (MS), 10 individuals with a neurological diagnosis of Parkinson's disease (PD), and 16 controls. All speakers took part in a larger project that included 17 speakers with MS, 12 speakers with PD, and 29 healthy controls (Tjaden, 2003; see also Tjaden & Wilding, 2005). Findings from these speech production studies indicated no difference in the strength of anticipatory vowel coarticulation for CV syllables produced by speakers with MS, speakers with PD, and age- and gender-matched healthy controls. As described next, stimuli produced by a subset of the larger speaker pool were selected for inclusion in the current study because acoustic measures of anticipatory vowel coarticulation for these speakers' utterances met certain operational criteria. The greater number of healthy talkers who produced stimuli included for study, as compared with speakers with MS or PD, also reflects criteria for stimulus selection.

Table 1 and Table 2 summarize speaker characteristics for individuals with MS and PD who produced stimuli included in the current study. Male and female speakers are denoted by speaker codes including the letter *M* or *F* (e.g., MSM1, PDF6). Dysarthria diagnoses reflect the consensus judgment of three speech-language pathologists (SLPs). These diagnoses were based on perceptual evaluation of vowel prolongation, diadochokinesis, the Grandfather Passage, and a brief conversational monologue. The three SLPs also judged overall dysarthria severity. Note that individuals with MS or PD who did not exhibit a perceptible dysarthria were not excluded, as instrumental measures of speech may reveal subperceptual characteristics of dysarthria. Intelligibility estimates in Tables 1 and 2 are reported for the purpose of characterizing the connected speech of individuals with MS and PD. As discussed in Tjaden (2003), these measures are magnitude estimates of intelligibility for the Grandfather Passage that were obtained using a fixed-modulus paradigm. Values reflect the geometric mean for 5 listeners. None of these 5 listeners participated in the vowel identification or intelligibility tasks reported in the current study. Relatively higher scale values in Tables 1 and 2 indicate relatively better intelligibility. Speech characteristics for individuals with PD reflect performance when optimally medicated, as audio recording took place 1 hr after participants ingested

Table 1. Speaker characteristics for participants diagnosed with multiple sclerosis.

Subject code	Age	Years post-diagnosis	Dysarthria diagnosis	Dysarthria severity	Deviant perceptual characteristics	Scaled intelligibility
MSF1	60	12	Spastic	Moderate	Strain-strangled, slow rate, short phrases	29
MSF2	42	8	Spastic–ataxic	Mild	Low pitch, imprecise consonants, excess and equal stress	238
MSF3	33	5	Spastic	Moderate	Strain-strangled, slow rate, voice tremor	57
MSF8	25	5	Ataxic–spastic	Moderate	Excess and equal stress, slow rate, short phrases	112
MSF9	50	9	Ataxic	Moderate	Slow rate, imprecise consonants, irregular artic breakdown	61
MSM2	45	4	Ataxic	Moderate	Slow rate, monopitch, irregular artic breakdown	91
MSM4	60	20	No perceptible dysarthria			248
MSM5	62	5	Ataxic	Mild	Excess and equal stress, harsh, voice tremor	168
MSM6	47	2	Ataxic	Mild	Hyponasal, imprecise consonants, voice tremor	97
MSM7	48	21	Ataxic	Moderate	Hyponasal, monopitch, monoloud	74

Note. Perceptual characteristics reflect judgments of three speech pathologists. Scaled estimates of intelligibility were obtained for the Grandfather Passage. Values correspond to the geometric mean for 5 listeners. Relatively higher numbers indicate better intelligibility, and relatively lower numbers indicate poorer intelligibility. MSF = multiple sclerosis female; MSM = multiple sclerosis male.

Table 2. Speaker characteristics for participants diagnosed with Parkinson's disease.

Subject code	Age	Years post-diagnosis	Dysarthria diagnosis	Dysarthria severity	Deviant perceptual characteristics	Scaled intelligibility
PDF1	42	6	Hypokinetic	Moderate	Monoloud, reduced loudness, variable rate	146
PDF3	50	3	Hypokinetic	Moderate/severe	Hypernasal, imprecise consonants, short rushes	38
PDF4	72	9	Hypokinetic	Moderate	Reduced loudness, variable rate, short rushes	115
PDF5	81	3	Hypokinetic	Severe	Repeated phonemes, low pitch, reduced loudness	27
PDF6	45	13	Hypokinetic	Moderate/severe	Fast rate, breathy voice, monoloud	95
PDM1	69	12	Hypokinetic	Moderate	Monopitch, monoloud, reduced stress	168
PDM3	72	4	Hyperkinetic	Mild	Harsh, forced inspiration/expiration, low pitch	178
PDM4	64	17	Hypokinetic	Moderate	Monopitch, monoloud, short rushes	78
PDM5	60	8	Hypokinetic	Moderate	Breathy, short rushes, repeated phonemes	70
PDM6	64	8	Hypokinetic/Hyperkinetic	Mild/moderate	Breathy, fast rate, voice stoppages	81

Note. Perceptual characteristics reflect judgments of three speech pathologists. Scaled estimates of intelligibility were obtained for the Grandfather Passage. Values reported correspond to the geometric mean for 5 listeners. Relatively higher numbers indicate better intelligibility, and relatively lower numbers indicate poorer intelligibility. PDF = Parkinson's disease female; PDM = Parkinson's disease male.

anti-Parkinsonian medications. Speakers with MS were recorded at times when individuals reportedly were well rested. Control speakers producing stimuli selected for inclusion in the current study included 8 women and 8 men ranging in age from 20 to 77 years ($M = 55$, $SD = 9$). (Speaker codes including the letter *M* or *F* refer to male and female speakers, respectively [e.g., CM12].)

All speakers scored at least 25/30 on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) and had pure-tone thresholds of 40 dB or better in at least one ear at 1, 2, and 4 kHz. In addition, none of the participants reported difficulty hearing speech in a quiet environment or wore hearing aids. Talkers were monolingual speakers of American English, and were judged by the first author to speak with a dialect characteristic of western New York state. Speakers were paid a modest participation fee.

Speech sample and recording procedures. Speakers produced CVC target words in the carrier phrase "It's a _____ again." The initial consonant of target words was /t/, /k/, or /s/, and the vowel was /i/ or /u/, yielding six different words. The final consonant typically was /d/, yielding three pairs of target words with the same initial consonant followed by the front vowel /i/ or the back vowel /u/ (i.e., *teed–tune, keyed–cooed, seed–sued*). The word *tune* was selected so that word-final consonants had similar place and voicing characteristics, and all target words were real words. Speakers produced three repetitions of each phrase in a different random order in his or her typical or habitual manner. Stimuli were recorded in a sound-treated booth using CSpeechSp (Milenkovic, 1997). The acoustic signal was transduced with an AKG C410 microphone positioned 9.5 cm and at a 50° angle from the center of the lips. The signal was preamplified, low pass filtered at 9.8 kHz, and digitized to a personal computer at 22 kHz using an analog to digital converter with 15-bit resolution.

For each participant, a 1 kHz calibration tone was recorded and saved to a computer disk for use in calculating intensity from the acoustic signal. The 1 kHz calibration tone was output from a function generator and played through a loudspeaker. The microphone of the sound level meter and the AKG C410 microphone used for recording speech samples were positioned parallel to each other at a distance of 9.5 cm from the loudspeaker, which corresponds to the mouth-to-microphone distance used to record speech samples. The 90 dB calibration tone, as verified by the sound level meter, subsequently was transduced by the AKG C410 microphone using the same preamplification and filter specifications used for recording of speech samples.

Acoustic index of coarticulation. First moment coefficient (M1) ratios were used to infer anticipatory vowel effects during the prevocalic consonant of the CVC target words. These types of ratios have been used to infer anticipatory vowel effects during a prevocalic consonant in other studies, including those investigating perception of anticipatory vowel coarticulation (e.g., Katz et al., 1991; Nittrouer, Studdert-Kennedy, & McGowan, 1989). Using CSpeech, first moments for fricatives were obtained for a 20-ms interval beginning 70 ms prior to vowel onset, and first moments for /t/ and /k/ were obtained for a 20-ms interval beginning at the burst release (see also Nittrouer, 1995; Sereno et al., 1987). For each speaker in the larger speaker pool reported in Tjaden (2003), a ratio was calculated for all unique pairs of /i/ versus /u/ vowel environments holding consonant context constant. Ratios approximating 1.0 indicate minimal differentiation of front versus back vowels during the prevocalic consonant, as indexed by first moment coefficients (e.g., first moment for CM12's Repetition 1 of *teed* = 6.502 kHz, first moment for CM12's Repetition 1 of *tune* = 6.497 kHz yields a ratio of 1.001). In contrast, ratios relatively greater than 1.0 suggest somewhat stronger anticipation of the upcoming vowel during the prevocalic consonant (e.g., first moment for MSM6's Repetition 1 of *seed* = 6.834 kHz, first moment for MSM6's Repetition 2 of *sued* = 5.237 yields a ratio of 1.30).

M1 ratios were inspected to identify pairs of target words for each consonant context (i.e., *seed–sued, teed–tune, keyed–cooed*) produced by a male and female speaker in each group characterized by an operationally defined *normal, under,* or *over* amount of coarticulation. *Normal coarticulation* was operationally defined as a ratio for a pair of target words approximating the mean ratio for the larger, appropriate gender-matched control group. For example, the mean M1 ratio for *seed–sued* target word pairings produced by the 16 control women reported in Tjaden (2003) was 1.07. CF4, MSF1, and PDF4 all produced *seed–sued* target word pairings with M1 ratios approximating 1.07 (i.e., CF4 = 1.07, MSF1 = 1.06, PDF4 = 1.08). *Undercoarticulation* was operationally defined as an M1 ratio approximating 1.00 for a given stimulus pairing, indicating minimal difference in first moments for consonants followed by front (/i/) versus back (/u/) vowels. *Overcoarticulation* was defined as the largest possible M1 ratio for a given pair of target words that was similar for male or female speakers in each of the three groups. For example, the largest M1 ratio for the *teed–tune* target word pairing that could be identified for male speakers approximated 1.39. CM13 and MSM5 both produced a *teed–tune* target word pairing with an M1 ratio of 1.39, and a *teed–tune* pairing with a similar ratio was identified for PDM4 (M1 ratio = 1.42).

For each consonant context, different stimuli produced by a male speaker and a female speaker from each group were identified that met the criteria for normal, under-, or overcoarticulation—with one exception. That is, stimuli meeting the operational definition of overcoarticulation for a female speaker with PD could not be identified for the *keyed–cooed* target word pairing. Thus, 54 different stimuli were identified for male speakers (i.e., 3 consonants × 2 vowels × 3 speaker groups × 3 amounts of coarticulation = 54), and 52 different stimuli were identified for female speakers for a grand total of 106 (54 + 52 = 106) experimental stimuli.

A two-way analysis of variance (ANOVA) was used to verify that M1 ratios differed for the operationally defined levels of coarticulation (normal, under, and over) and to identify any differences in M1 ratios among speaker groups (control, MS, and PD). Gender was used as a covariate, and separate ANOVAs were performed for each consonant context. The fixed main effect of coarticulation level was significant for /s/, $F(2, 8) = 91.90$, $p < .01$, with post hoc Tukey tests indicating significantly different ratios for all pairwise comparisons ($p < .01$). The main effect of coarticulation level was also significant for /t/ and /k/: /t/, $F(2, 8) = 67.95$, $p < .01$; /k/, $F(2, 7) = 37.33$, $p < .01$. Post hoc tests for /t/ and /k/ indicated that ratios for all pairwise comparisons, except the normal–under contrast, were significantly different ($p < .01$). The group main effect was not significant for any of the ANOVAs. Thus, consistent with our earlier speech production study indicating no group differences in the strength of anticipatory vowel coarticulation for the larger pool of speakers with PD, MS, and healthy controls (Tjaden, 2003), stimulus selection yielded M1 ratios suggestive of similar amounts of anticipatory vowel coarticulation for all speaker groups. Stimulus selection also yielded M1 ratios suggestive of different amounts of anticipatory vowel coarticulation for the operationally defined normal, under, and over categories, although M1 ratios for the normal and under categories were not statistically different for /t/ and /k/ contexts. This finding is considered further in the discussion.

Supplemental acoustic measures. Additional acoustic measures obtained for the experimental stimuli included the temporal midpoint first moment coefficient for /s/, root-mean-square (RMS) intensity of the closure interval for /t/ and /k/, vowel F2 midpoint, and duration of the prevocalic consonant. Supplemental acoustic measures were used as covariates in the data analyses to evaluate the influence of signal characteristics other than M1 ratios or strength of anticipatory vowel coarticulation on perceptual judgments. Measures are discussed next.

Poor production of consonants and/or vowels could interfere with listeners perceiving anticipatory vowel information during a prevocalic consonant. Degraded segmental articulation also could impact intelligibility for the entire target word. The first moment coefficient at the temporal midpoint of /s/ provides an index of the articulatory integrity that is more or less unaffected by anticipatory or perseveratory coarticulation. Using TF32, the time–frequency analysis software program for 32-bit Microsoft Windows, the first moment for each /s/ was obtained for a 20-ms interval centered at the temporal midpoint of the fricative. A similar measure for /t/ and /k/ is unavailable because anticipatory vowel effects for these consonants are quantified for a temporal interval beginning at the release burst. Energy during the closure interval for /t/ and /k/ provides a gross indication of the integrity of stop consonant production, however, with relatively larger intensity measures suggesting relatively poorer consonant precision (Ackermann & Ziegler, 1991). Thus, the RMS intensity of the stop consonant closure interval was obtained using TF32. This measure was converted to dB SPL in Microsoft Excel, with reference to the calibration tone recorded for each speaker. Vowel F2 midpoint measures were used to index articulatory integrity of vowels. The entire vowel was first identified from the combined waveform and wideband (300–400 Hz) spectrographic displays of TF32. Vowel onset was taken as the first glottal pulse following frication or the release burst. Vowel offset was taken as the last glottal pulse prior to the closure interval for word-final /d/ or the last high-intensity glottal pulse prior to the reduction in intensity for the /n/ in *tune*. F2 values were obtained at the midpoint using the combined wideband spectrographic and linear predictive coding displays of TF32.

Stimulus duration could also affect perceptual judgments. This was of particular concern in the vowel identification task in which listeners were presented with the aperiodic portion of prevocalic consonants for identification of the upcoming vowel. For /s/, consonant duration was taken as the temporal interval between fricative onset and offset. Frication onset was defined as the beginning of frication following the last glottal pulse for the word *a* in the carrier phrase. Frication offset was defined as the end of frication just prior to the first glottal pulse for the following vowel. Consonant duration for /t/ and /k/ was operationally defined as the temporal interval between the left edge of the release burst and the onset of the following vowel. This interval corresponds to the aperiodic portion of /t/ and /k/, and is typically termed *voice onset time* (VOT). The term *consonant duration* is used in the present study so that the same terminology can be used to indicate the aperiodic portion of both fricatives and stops. Duration measures were obtained using the combined waveform and wideband spectrographic displays.

Vowel Identification Task: Research Questions 1 and 2

Method

Listeners. Two men and 8 women ranging in age from 18 to 55 years ($M = 27$, $SD = 13$) served as listeners. All listeners were right-handed, were native speakers of American English, had no training in phonetics or communication disorders and sciences, and had bilateral pure-tone thresholds of no more than 20 dB at octave frequencies between 250 Hz and 8000 Hz. Listeners were recruited from the student and staff population at the University at Buffalo and were paid a modest participation fee.

Stimuli preparation and procedures. Using the combined waveform and RMS displays of TF32, the entire aperiodic portion of the prevocalic consonant of target words was identified and saved to a computer disk. This is the segment presented to listeners in other studies investigating perception of anticipatory vowel information (e.g., Baum, 1998; Katz, 1988) and corresponds to the consonant duration interval discussed in the preceding section. Zero crossing points on the RMS display were used when extracting consonants to minimize noise at stimuli onsets and offsets. Prior to presentation to listeners, stimuli were equated for peak intensity using Cool-Edit Pro (Syntrillium Corporation) to control for any influence of audibility on vowel identification.

Stimuli were presented via headphones to the right ear at an average peak intensity of 70 dB SPL, as determined using a Quest (Model 155) sound level meter. Individual listeners completed the task in a double-walled, sound-treated booth, and stimulus presentation was controlled via a personal computer that also recorded listeners' responses. Trials consisted of two presentations of a given stimulus (i.e., /t/ in *teed* produced by a given speaker) separated by a 1-s pause. Written and verbal instructions directed listeners to identify the following vowel as either *ee* or *oo* by pressing the appropriately labeled button on a response box after hearing the second presentation of the stimulus. A listener response prompted the computer to deliver the next trial after a 1-s pause. Five trials of each stimulus were presented in random order, blocked by consonant (/s/, /t/, or /k/) and gender for a total of 530 experimental trials (106 stimuli × 5 repetitions = 530). Blocking stimulus presentation by consonant and gender was performed to take into consideration the fact that first moment ratios for a given level of coarticulation differed as a function of consonant context and gender.

Each of the six experimental blocks (3 consonants × 2 genders) was preceded by 6 practice trials. A practice session consisting of all 36 practice trials preceded the beginning of the experiment. Practice stimuli differed from those used in the experiment. Feedback concerning overall accuracy of response was provided in the form of a percent correct score at the end of each practice task as well as at the end of each experimental block. Two random orderings of the stimuli were generated, and 5 listeners were randomly assigned to each ordering. The vowel identification task took approximately 1 hr per listener.

Following studies investigating perception of coarticulatory information in aphasia (e.g., Baum, 1998), the initial data analysis examined whether response accuracy was above chance level. For each consonant context and speaker group, listener responses were pooled across level of coarticulation and gender, and binomial probabilities were computed. Given two response choices, a chance level of response corresponds to .50 or 50% accuracy. For the main data analysis, an overall percent correct score was calculated for each of the 106 stimuli by pooling the 10 listeners' responses. Overall percent correct scores were arcsine transformed, and for each consonant context an ANOVA was used to determine differences in percent correct scores as a function of speaker group (control, MS, and PD) and coarticulation level (normal, over, and under).

Vowel type (/i/, /u/) was also included as a factor in the ANOVAs, given the possibility of differences in identification accuracy for /i/ and /u/ (e.g., Baum, 1998; Yeni-Komshian & Soli, 1981). Group, coarticulation, and vowel were treated as fixed effects, and gender was included as a covariate. The three ANOVAs were repeated using the same fixed main effects, but covariates included supplemental acoustic measures as well as gender. Similar findings for the two sets of ANOVAs would suggest that signal characteristics such as integrity of vowel and consonant production, as well as consonant duration, did not influence perceptual judgments. The Tukey–Kramer adjustment for multiple comparisons was used for post hoc testing. An alpha level of .05 was used for all analyses. This was deemed reasonable for an initial study investigating perception of coarticulatory information in dysarthria.

Results

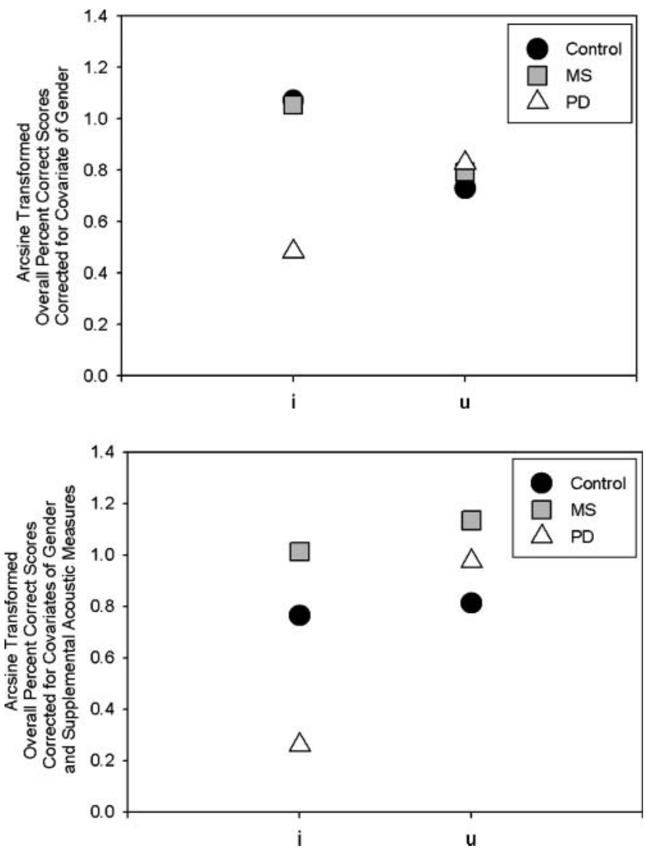
Vowel identification results. Table 3 reports raw percent correct scores. For example, the percent correct score of 99% for /i/ in *keyed* produced by control speakers reflects the finding that listeners correctly identified the vowel in 298 of 300 trials (e.g., 2 genders × 3 levels of coarticulation × 5 stimulus repetitions × 10 listeners = 300 trials). Asterisks designate stimuli for which vowels were accurately identified at a level greater than chance (binomial distributions, $p < .01$). Table 3 shows that with the exception of *seed* produced by speakers with PD for which the vowel was correctly identified only 46% of the time, listeners consistently identified vowels with greater than chance accuracy when presented with only the aperiodic portion of the prevocalic consonant.

Table 3. Raw percent correct scores for vowel identification.

Vowel identification	Control speakers	Speakers with MS	Speakers with PD
/s/			
/i/	84*	83*	46
/u/	65*	69*	69*
Overall	76	77	57
/t/			
/i/	89*	93*	76*
/u/	88*	94*	89*
Overall	89	94	82
/k/			
/i/	99*	99*	96*
/u/	97*	98*	99*
Overall	98	99	98

Note. Rows labeled "Overall" indicate percent correct scores pooled across vowel contexts. Asterisks indicate stimuli for which vowels were correctly identified with greater than chance accuracy (binomial distribution, $p < .01$). Chance level of response corresponds to 50% accuracy.

Figure 1. This figure plots estimated marginal means for the arcsine-transformed percent correct scores for /s/ obtained in the vowel identification task. The upper panel reports data for analyses in which gender was the only covariate. The lower panel reports analyses in which gender and supplemental acoustic measures were included as covariates.



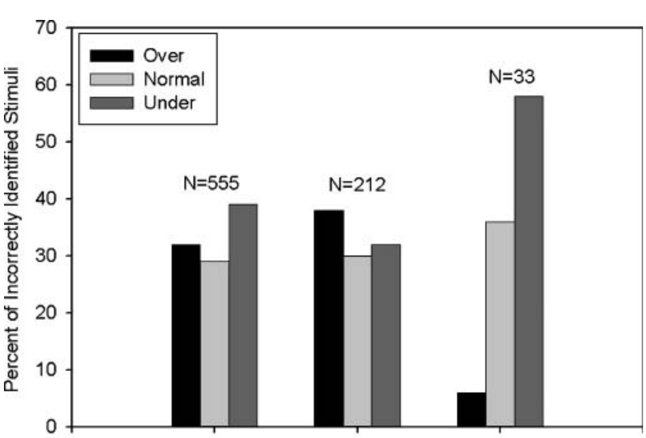
In the first ANOVA for /s/, for which gender was the only covariate, the Group × Vowel interaction was significant, $F(2, 17) = 5.43$, $p = .02$. The upper panel of Figure 1 illustrates this finding by plotting estimated marginal means for the arcsine-transformed percent correct scores. Post hoc tests indicated that identification accuracy for /i/ was reduced for the PD group; that is, MS versus PD, $t = -3.521$, $p = .03$; control versus PD, $t = -3.625$, $p = .02$. No other post hoc comparisons investigating the source of the interaction were significant. In the second ANOVA for /s/, wherein gender and supplemental acoustic measures were included as covariates, the main effect of group, $F(2, 14) = 4.107$, $p = .04$, and the Group × Vowel interaction, $F(4, 14) = 4.56$, $p = .03$, were significant. The lower panel of Figure 1 illustrates the interaction. As in the upper panel of Figure 1, post hoc tests confirmed that identification accuracy for /i/ differed for stimuli produced by the PD and MS groups, $t = -3.58$, $p = .03$. No other post hoc tests investigating the source of the interaction in the lower panel of Figure 1 were significant. Finally, none of the main effects or interactions were significant in the ANOVAs for /t/ or /k/.

Incorrectly identified stimuli were further inspected to determine if the pattern of vowel identification errors differed across levels of coarticulation. This analysis was undertaken as it was speculated that a ceiling effect for vowel identification accuracy might have masked trends of interest. Figure 2 reports the percentage of incorrect responses as a function of coarticulation level. The number of tokens used for computing percentages is also reported. Chi-square tests indicated that the proportion of incorrectly identified normal, under, and over tokens differed for /s/, $\chi^2(2, N = 555) = 8.832$, $p = .01$, and /k/, $\chi^2(2, N = 33) = 13.273$, $p = .001$, but not /t/. Figure 2 indicates that undercoarticulated stimuli made up the largest proportion of incorrectly identified tokens for both /s/ and /k/.

Summary of findings for vowel identification task. Research Question 1 asked whether listeners can use anticipatory vowel information in prevocalic consonants produced by speakers with dysarthria to identify the upcoming vowel. Results suggest that the answer to this question is yes. That is, response accuracy for stimuli produced by speakers with MS and PD was above chance level, and the ANOVAs indicated no difference in vowel identification accuracy for stimuli produced by the different speaker groups. The exception was /i/ in *seed* produced by speakers with PD, for which vowel identification accuracy was reduced relative to other speaker groups. The similar pattern of findings for both ANOVAs further suggests that this result was not wholly explained by factors such as consonant duration or a poorly articulated consonant or vowel influencing perceptual judgments. Data in the upper and lower panels of Figure 1 are not identical, however, indicating at least some influence of supplemental acoustic measures on perceptual judgments for the /s/ consonant context.

Research Question 2 asked whether listeners are sensitive to interspeaker variation in the strength of anticipatory vowel coarticulation during prevocalic consonants produced by speakers with dysarthria and/or healthy controls. Results suggest that the answer to this question is generally no—albeit a qualified no. That is, on the one hand, neither the main effect of coarticulation level nor the Group × Coarticulation Level interaction was significant in the ANOVAs, indicating similar identification accuracy for normal, under-, and overcoarticulated stimuli produced by all speaker groups. The fact that these results held for both sets of ANOVAs suggests that factors such as consonant duration or a poorly produced consonant or vowel did not strongly affect listeners' ability to detect interspeaker differences in anticipatory vowel coarticulation during a prevocalic consonant. On the other hand, Figure 2 shows that undercoarticulated stimuli tended to make up the largest proportion of incorrectly identified tokens.

Figure 2. Percentage of incorrectly identified vowels in the vowel identification task is reported as a function of strength of coarticulation. The total number of tokens used for computing percentages is also reported for each consonant context.



Intelligibility Task: Research Question 3

Method

Listeners. Listeners included 6 men and 4 women ranging in age from 19 to 38 years ($M = 25$, $SD = 6$) for a total of 10 participants. Inclusionary criteria, recruitment procedures, and participant remuneration were identical to those described for the vowel identification task.

Procedures. The entire carrier phrase (e.g., "It's a _____ again") was presented to listeners for identification of the target word. A multiple-choice task modeled after Yorkston and Beukelman's (1981) single-word intelligibility test was used, wherein listeners selected a target word from a list of 12 alternatives. The 12 response alternatives included the target word, the option "none of the above," and 10 response alternatives consisting of a quasirandom selection of CVC words taken from the larger stimulus pool reported in Tjaden (2003). This larger stimulus pool included words beginning with /t, k, s, ʃ/, followed by /i, ae, u, ɑ/, and ending in /d/.

Response alternatives always included at least one item beginning with the same word–initial consonant as the target word and at least one item containing the same vowel as the target word. The 106 experimental utterances and 54 foils were presented to listeners in a random order. Foil stimuli were taken from the larger speaker pool reported in Tjaden (2003), and did not include items produced by speakers in the current study.

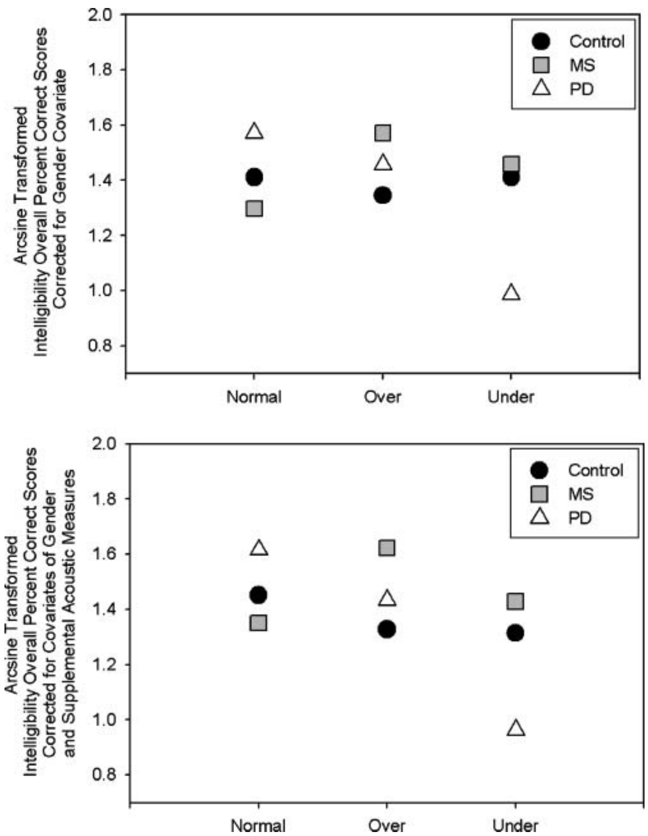
Stimuli were presented via headphones to listeners' right ears at an average peak intensity of 70 dB SPL in a double-walled, sound-treated booth. Stimulus presentation was controlled via personal computer. Listeners were provided with a response sheet, and were given written and verbal instructions. Listeners were instructed that they would be hearing a series of phrases containing a target word, and their task was to circle the word they heard on the response sheet. Listeners were further instructed that they would hear the words *seed, sued, sad, sod, she'd, shooed, shad, shod, teed, tune, tad, todd, keyed, cooed, cad, cod*. Trials consisted of two presentations of a given target phrase separated by a 1-s pause. Listeners were given unlimited time to respond, and were instructed to press a button on a response box to prompt the computer to deliver the next trial. Six practice trials preceded the experiment, and the entire listening task lasted approximately 45 min per listener.

Data analyses were modeled after those for the vowel identification task. For each of the 106 stimuli, listener responses were pooled, and an overall percent correct score was computed. Overall percent correct scores were arcsine transformed and for each consonant context, an ANOVA was used to determine differences in percent correct scores as a function of group and coarticulation level. Vowel was also included as a fixed main effect in the ANOVAs. Gender was used as a covariate in the first set of ANOVAs. Gender and supplemental acoustic measures were used as covariates in the second set of ANOVAs. Similar findings for both sets of ANOVAs would indicate that signal characteristics, such as consonant duration and articulatory integrity of vowels and consonants, did not influence perceptual judgments.

Results

Intelligibility results. The Group × Coarticulation Level interaction, $F(4, 17) = 3.411$, $p = .03$, was significant in the first ANOVA for /s/, for which gender was the only covariate. This finding is illustrated in the upper panel of Figure 3. Perhaps the most obvious trend is for intelligibility to be reduced for undercoarticulated stimuli produced by speakers with PD. Significant post hoc comparisons evaluating the source of the interaction included the PD normal–PD under contrast, $t = –3.578$, $p = .04$, and the MS over–PD under contrast, $t = –3.578$, $p = .04$. No other post hoc comparisons were significant. The vowel main effect was also significant in the first ANOVA for /s/, with better intelligibility for *seed* versus *sued*, $F(2, 17) = 9.794$, $p = .01$. Neither the Group × Coarticulation Level interaction nor the vowel main effect was significant in the second ANOVA for /s/, wherein gender and supplemental acoustic measures were included as covariates. For comparison purposes, however, the lower panel of Figure 3 reports average, transformed percent correct scores—corrected for covariates of gender and supplemental acoustic measures. The pattern of results is remarkably similar to that shown in the upper panel of Figure 3. Finally, in the ANOVAs for /t/ and /k/, no main effects or interactions were significant.

Figure 3. Estimated marginal means for the intelligibility task are reported for /s/ words. Measures reported in the upper panel reflect adjustments for the covariate of gender. Data reported in the lower panel reflect adjustments for the covariates of gender and supplemental acoustic measures.



Summary of findings for intelligibility task. Research Question 3 asked whether interspeaker variation in anticipatory vowel coarticulation for neurologically normal speakers and/or speakers with dysarthria is reflected in measures of intelligibility. The majority of findings suggest that the answer to this question is no. That is, intelligibility for stimuli produced by speakers with dysarthria and healthy controls did not typically differ as a function of level of coarticulation.

Undercoarticulated /s/ words produced by speakers with PD were relatively less intelligible, however. As suggested by the similar results in both panels of Figure 3, the trend toward reduced intelligibility for undercoarticulated stimuli produced by the PD group appears to be related to coarticulatory characteristics of the stimuli, although articulatory characteristics of the word–initial consonant or vowel as well as consonant duration influenced perceptual judgments to some extent as indicated by the finding that the Group × Coarticulation Level interaction only approached significance ($p = .054$) when supplemental acoustic measures were included as covariates. Relatedly, the finding of statistically better intelligibility for *seed* versus *sued* in only the first ANOVA for /s/ words suggests that segmental integrity of the word–initial consonant or vowel as well as consonant duration contributed to the intelligibility difference for *seed* and *sued*. Finally, the fact that the Group × Vowel interaction for /s/ words was not significant indicates that the reduced identification accuracy for /i/ in *seed* produced by speakers with PD reported in the vowel identification task was not associated with reduced intelligibility for the entire word. Intelligibility for *seed*—collapsed across levels of coarticulation and gender—was 97%, 98%, and 100% for the PD, MS, and control groups, respectively.

Discussion

The present study sought to further investigate the perceptual relevance of intrasyllabic, anticipatory vowel coarticulation. The issue of how and when coarticulatory cues are important for perception is important for theories of speech perception and spoken word recognition in which listeners exploit coarticulatory information in the acoustic speech stream. Studies that add to our understanding of factors that may impact intelligibility are also clinically relevant.

For the most part, listeners correctly identified vowels in CVC words produced by speakers with PD, speakers with MS, and healthy controls when presented with only the aperiodic portion of the prevocalic stop or fricative. This finding suggests that listeners are tuned to similar types of information in the acoustic speech stream, irrespective of the source or speaker (see discussion in Weismer & Martin, 1992, regarding the use of similar perceptual strategies for perceiving the acoustic signal of dysarthria and neurologically normal speech). Theories and models developed to account for perception of neurologically normal speech therefore might be extended to dysarthria, although studies including stimuli produced by individuals with more severe dysarthria would provide stronger support for this suggestion.

These statements must be tempered by the finding that vowel identification accuracy was below chance level for /i/ in *seed* produced by speakers with PD. Studies investigating perception of vowel coarticulatory information in prevocalic consonants produced by children with hearing impairment have also reported poorer identification accuracy for /i/ compared with /u/ (Waldstein & Baum, 1994). These investigators hypothesized that identification accuracy for /u/ from a prevocalic consonant was relatively better because children with hearing impairment learn to produce better /u/ exemplars, which can be attributed to the highly visible lip rounding. It was also speculated that speakers with hearing impairment may have difficulty producing fronted consonants in the context of front vowels, and consonant imprecision may therefore interfere with perception of vowel coarticulatory information during a prevocalic consonant. Lingual function, as compared with lip or jaw function, is disproportionately affected in dysarthria. Thus, it might be speculated that the PD speakers had difficulty producing good exemplars of /i/ in the context /s/ because the midline tongue groove for /s/ demands such a high degree of lingual precision that the tongue could not adequately adjust for the upcoming vowel. Although articulatory imprecision for /s/ could also have influenced perceptual judgments, speakers are likely aware of the importance of word–initial consonants for lexical access and maximize articulatory precision for these phonetic events.

When vowel midpoint F2 measures and first moment coefficients were included as covariates in data analyses for the vowel identification task, however, identification accuracy for /i/ in *seed* from the prevocalic /s/ was still reduced for stimuli produced by speakers with PD relative to other speaker groups. This result suggests that vowel or consonant articulatory imprecision did not explain the reduced identification accuracy.

Thus, the source of the reduced identification accuracy for /i/ in *seed* produced by the PD group is unclear. As noted in the Limitations section, however, vowel midpoint F2 measures and consonant first moment coefficients only index segmental articulatory adequacy at a given point in time during phonetic events. The possibility that the reduced identification accuracy of /i/ in *seed* produced by the PD group was related to a poorly produced vowel or consonant, therefore, cannot be entirely ruled out.

Difficulty in perceiving anticipatory vowel information during a prevocalic /s/ produced by speakers with PD had little impact on intelligibility, however. Intelligibility of *seed* produced by speakers with PD was similar to intelligibility of *seed* produced by other speaker groups, even though listeners could not identify the vowel in PD stimuli with better than chance accuracy when presented with only the prevocalic fricative. The implication is that information that is meaningful in a simple vowel identification task may not necessarily be indicative of the importance of that information for recovering the speaker's intended message. This suggestion obviously requires support from studies including a larger speaker and stimulus pool. If additional studies report similar findings, however, this may suggest that coarticulatory deviancies noted for populations like AOS are not strongly related to intelligibility deficits for these speakers. Taking this idea a step further, behavioral therapies aimed at remediating coarticulatory deviancies that have been reported in some speakers with AOS may have little bearing on intelligibility (see Southwood et al., 1997, for related discussion).

Interspeaker differences in level of coarticulation were generally not perceptually relevant, as indicated by the fact that vowel identification accuracy and intelligibility typically did not differ for normal, under-, and overcoarticulated stimuli. This result might be taken to suggest that listeners ignore or normalize for interspeaker variation in coarticulation, and as such, perceptual theories need not be concerned with the type of interspeaker coarticulatory variation studied in the current project. On the other hand, studies have found that listeners are less accurate identifying vowels from only a prevocalic fricative for CV stimuli characterized by relatively lower anticipatory vowel coarticulation (Nittrouer & Whalen, 1989; Yeni-Komshian & Soli, 1981).

Relatedy, it is worth asking how listeners in the current study were even able to identify vowels in undercoarticulated stimuli, given that these stimuli were characterized by M1 ratios approximating 1.0. Both points may be related to the manner in which coarticulation level was quantified in the current study as well as the consonant interval presented to listeners in the vowel identification task.

Nittrouer and Whalen (1989) used a gating paradigm to study perception of anticipatory vowel information, and Yeni-Komshian and Soli (1981) studied anticipatory vowel effects across the entire frication interval. Thus, both of these studies were interested in the time course of coarticulation. Coarticulatory effects in the present study were measured at a point at which anticipatory vowel effects on stop consonants and fricatives have been observed (e.g., Nittrouer, 1995; Sereno et al., 1987). As discussed in Waldstein and Baum (1994), a gating paradigm is not appropriate for stop consonant + vowel stimuli, and listeners in the current study, therefore, were presented with the entire aperiodic portion of the consonant in the vowel identification task. As vowel onset is approached in the consonant of a CV syllable, vowel coarticulatory effects present at earlier points tend to strengthen (Nittrouer & Whalen, 1989), or if coarticulatory effects were absent earlier in the consonant, they become apparent (Southwood et al., 1997; Yeni-Komshian & Soli, 1981). Listeners were exposed to anticipatory vowel information throughout the entire consonant interval in the vowel identification task, and this likely explains the high vowel identification accuracy for undercoarticulated stimuli.

To address this methodological issue, a second group of 10 listeners was presented with only the 20-ms portion of consonants from which first moment coefficients were obtained. Procedures were identical to those described for the vowel identification task. In addition to providing a closer correspondence between the acoustic measures of coarticulation and the actual speech sample heard by listeners, it was expected that the shorter stimuli would make the perceptual task more difficult and thus more sensitive to listener effects (Nittrouer & Whalen, 1989). Vowel identification accuracy for all speaker groups was only above chance for /ki/ and /ku/, as indicated by binomial probabilities (control: /ki/ = 86%, /ku/ = 77%; MS: /ki/ = 78%, /ku/ = 86%; PD: /ki/ = 75%, /ku/ = 76%). The arcsine-transformed overall percent correct scores for /k/ were evaluated using data analysis procedures described for the original vowel identification task, with the exception that consonant duration was not included as a covariate in the second ANOVA. Similar to the original findings, no main effects or interactions were significant. Thus, listeners were not less accurate identifying vowels for stimuli characterized by relatively lower coarticulation, even when the stimulus interval heard by listeners corresponded exactly to the interval over which coarticulatory effects were measured—a finding consistent with the conclusion that the type of interspeaker coarticulatory variability of interest in the current project is not perceptually important. The supplemental analysis of incorrect responses for the vowel identification task and the trend toward reduced intelligibility of undercoarticulated /s/ words produced by the PD group suggest that listeners were at least marginally sensitive to interspeaker differences in coarticulation, however. Taken together, the available evidence seems to suggest that the perceptual effects of interspeaker variations in coarticulation are subtle and may only be revealed in some instances.

Finally, the influence of signal characteristics other than M1 ratios on perceptual judgments deserves additional consideration. Two sets of data analyses were performed to explore the potential effects of consonant duration as well as vowel and consonant articulatory adequacy on perceptual judgments. For the most part, the pattern of findings for the two analyses was similar—both for the vowel identification task and the intelligibility task—suggesting that supplemental acoustic measures did not impact perceptual judgments. Comparison of data in the upper and lower panels of Figure 1, however, indicates that signal characteristics other than M1 ratios had some degree of impact on vowel identification accuracy for /s/ stimuli—even though the overall pattern of results from the statistical analysis was similar for both panels in Figure 1 (i.e., significant Group × Vowel interaction in both panels of Figure 1, which is attributed to reduced identification accuracy for /i/ in *seed* produced by the PD group). Findings from the intelligibility task also indicated subtle effects of supplemental acoustic measures on perceptual judgments, at least for /s/ words. Thus, it appears that a variety of cues or information sources in the acoustic signal have the potential to influence perception of coarticulatory information and that these factors need to be considered in studies investigating the perceptual relevance of coarticulation—even studies of neurologically normal speech.

Limitations and Caveats

A variety of factors suggest cautious interpretation of results as well as variables to consider in future studies investigating perception of coarticulatory information. Intelligibility of target words did not typically differ among speaker groups—a finding indicating that speakers with MS and PD had mild dysarthria. It may be the case that coarticulatory information in the acoustic signal of persons with more severe dysarthria—even those with intact coarticulatory patterns—is not perceptible, which contributes to reductions in intelligibility.

As previously noted, the acoustic measures used to index articulatory adequacy of fricatives and vowels were taken at a single point in time or over a brief temporal interval. These kinds of "slice-in-time" measures do not capture articulatory characteristics of the entire phonetic segment, such as the RMS intensity measure used to index articulatory adequacy for stops that was calculated over the entire consonant closure interval. However, RMS intensity of stop consonant closure intervals is problematic as a measure of articulatory adequacy in so far as some neurologically normal speakers are known to spirantize dorsal stops (e.g., Kent & Moll, 1972; Kuhnert & Hoole, 2004). Future studies may wish to consider using kinematic indices of segmental articulatory adequacy.

The fact that M1 ratios for /t/ and /k/ contexts were not statistically different for the normal and undercoarticulated categories may have contributed to the finding that listeners were not sensitive to interspeaker variation in level of coarticulation. Studies using synthesized speech materials would be one way to ensure precise differences in level of coarticulation among stimuli. As noted by Katz (2000), however, these kinds of laboratory tasks may indicate what listeners can do but not what listeners actually do when perceiving naturally produced speech. Finally, M1 ratios—the acoustic measure of coarticulation—were calculated for pairs of stimuli, but perceptual tasks required listeners to judge single stimuli. It might be argued that perceptual tasks involving paired comparisons would provide a closer correspondence to the speech production measure used to quantify coarticulation. Our approach of quantifying strength of anticipatory vowel coarticulation for pairs of stimuli, but assessing the perceptual relevance of coarticulatory information for individual stimuli, follows from production–perception studies investigating coarticulation in aphasia (Baum, 1998; Katz, 1988). It could be argued that listeners might have been able to detect subtle differences in coarticulation using a discrimination paradigm, but that such ability might not change how a listener labels or identifies the target sounds. Studies comparing findings for a paired-comparison perceptual task and a perceptual task involving judgments of single stimuli help to determine if a given type of perceptual task is preferred for studying perception of coarticulatory information.

Acknowledgments

Portions of this study were presented at the Fall 2003 and Spring 2004 Meeting of the Acoustical Society of America. Many thanks to members of the Speech Editoral group for their patience and constructive comments during the review process. Research was supported by National Institute on Deafness and Other Communication Diseases Grant R01 DC04689.

References

Ackermann, H., & Ziegler, W. (1991). Articulatory deficits in Parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry, 54,* 1093–1098.

Baum, S. R. (1998). Anticipatory coarticulation in aphasia: Effects of utterance complexity. *Brain and Language, 63,* 357–380.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30,* 591–627.

Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech waveform. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processing* (pp. 360–385). Hillsdale, NJ: Erlbaum.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning in a distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23,* 439–432.

Goldinger, S. D., Pisoni, D. B., & Luce, P. A. (1996). Speech perception and spoken word recognition: Research and theory. In N. J. Lass (Ed.), *Principles of experimental phonetics* (pp. 277–327). St. Louis, MO: Mosby-Year Book.

Hertrich, I., & Ackermann, H. (1999). Temporal and spectral aspects of coarticulation in ataxic dysarthria:

An acoustic analysis. *Journal of Speech, Language, and Hearing Research, 42,* 367–381.

Katz, W. F. (1988). Anticipatory coarticulation in aphasia: Acoustic and perceptual data. *Brain and Language, 35,* 340–368.

Katz, W. F. (2000). Anticipatory coarticulation and aphasia: Implications for phonetic theories. *Journal of Phonetics, 28,* 313–334.

Katz, W. F., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech and Hearing Research, 34,* 1222–1232.

Kent, R., & Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research, 15,* 453–473.

Kuehn, D. P., & Moll, K. L. (1972). Perceptual effects of forward coarticulation. *Journal of Speech and Hearing Research, 15,* 654–664.

Kuhnert, B., & Hoole, P. (2004). Speaker-specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics and Phonetics, 18,* 559–575.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489–494.

Lubker, J. F., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological and linguistic. *Journal of the Acoustical Society of America*, *71,* 437–448.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62,* 615–625.

Luce, P., & McClellan, C. (2005). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. B. Remez (Eds.), *The handbook of speech perception* (pp. 591–609). Malden, MA: Blackwell.

Manuel, S. Y. (1987). *Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages*. Unpublished dissertation, Yale University.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America, 69,* 559–567.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel–stop consonant–vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 473–488.

Mattys, S. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 397–408.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McQueen, J., & Cutler, A. (1997). Cognitive processes in speech perception. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 566–586). Oxford, England: Blackwell.

Milenkovic, P. (1997). CSpeechSp 4.0 [Computer software]. Madison, WI: Department of Electrical and Computer Engineering, University of Wisconsin.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America, 97,* 520–530.

Nittrouer, S., & Whalen, D. H. (1989). The perceptual effects of child–adult differences in fricative–vowel coarticulation. *Journal of the Acoustical Society of America, 86,* 1266–1276.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative–vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120–132.

Ostreicher, H. J., & Sharf, D. J. (1976). Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics, 4,* 285–301.

Sereno, J. A., Baum, S. R., Marean, G. C., & Lieberman, P. (1987). Acoustic analysis and perceptual data on anticipatory labial coarticulation in adults and children. *Journal of the Acoustical Society of America, 81,* 512–519.

Southwood, H., Dagenais, P. A., Sutphin, S. M., & Garcia, J. M. (1997). Coarticulation in apraxia of speech: A perceptual, acoustic, and electropalatographic study. *Clinical Linguistics and Phonetics, 11,* 179–203.

Sussman, H. M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: The orderly output constraint. *Behavioral and Brain Sciences, 21,* 241–299.

Tjaden, K. (2003). Anticipatory coarticulation in multiple sclerosis and Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 46,* 990–1008.

Tjaden, K., & Turner, G. (1997). Spectral properties of fricatives in ALS. *Journal of Speech, Language, and Hearing Research, 40,* 1358–1372.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783.

Tjaden, K., & Wilding, G. E. (2005). Effect of rate reduction and increased loudness on acoustic measures of anticipatory coarticulation in multiple sclerosis and Parkinson disease. *Journal of Speech, Language, and Hearing Research, 48,* 261–277.

Waldstein, R. S., & Baum, S. R. (1994). Perception of coarticulatory cues in the speech of children with profound hearing loss and children with normal hearing. *Journal of Speech and Hearing Research, 37,* 952–959.

Weismer, G., & Martin, R. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 68–118). Amsterdam: John Benjamins.

Weismer, G., Yunusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria: An X-ray microbeam study. *Journal of Speech, Language, and Hearing Research, 46,* 1247–1261.

Winitz, H., Scheib, M., & Reeds, J. (1972). Identification of stops and vowels from the burst portion of /p t k/ isolated from conversational speech. *Journal of the Acoustical Society of America, 51,* 1309–1317.

Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative–vowel coarticulation. *Journal of the Acoustical Society of America, 70,* 966–975.

Yorkston, K., & Beukelman, D. (1981). *Assessment of the intelligibility of dysarthric speech*. Tigard, OR: CC Publications.

Ziegler, W., & von Cramon, D. R. (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language, 26,* 117–130.

Ziegler, W., & von Cramon, D. R. (1986). Spastic dysarthria after acquired brain injury: An acoustic study. *British Journal of Disorders of Communication, 21,* 173–187.

Received October 25, 2004

Revision received March 31, 2005

Accepted December 15, 2005

DOI: 10.1044/1092-4388(2006/064)

Contact author: Kris Tjaden, Department of Communicative Disorders and Sciences, University at Buffalo, 122 Cary Hall, 3435 Main Street, Buffalo, NY 14214.
E-mail: tjaden@acsu.buffalo.edu