Published in final edited form as:

Speech Commun. 2011 July ; 53(6): 842–854. doi:10.1016/j.specom.2011.02.003.

Detecting anticipatory effects in speech articulation by means of spectral coefficient analyses

Yongqiang Feng^a, Grace J. Hao^b, Steve A. Xue^c, and Ludo Max^{d,e,*}

^aChinese Academy of Sciences, Institute of Acoustics, 21 North 4th Ring West Road, Beijing 100190, China

^bNorth Carolina Central University, Department of Communication Disorders, Durham, NC 27707, USA

^cUniversity of Hong Kong, Division of Speech and Hearing Sciences, 34 Hospital Road, Hong Kong

^dUniversity of Washington, Department of Speech and Hearing Sciences, 1417 N.E. 42nd Street, Seattle, WA 98105-6246, USA

^eHaskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511, USA

Abstract

Few acoustic studies have attempted to examine anticipatory effects in the earliest part of the release of stop consonants. We investigated the ability of spectral coefficients to reveal anticipatory coarticulation in the burst and early aspiration of stops in monosyllables. Twenty American English speakers produced stop (/k,t,p/) – vowel (/æ,i,o/) – stop (/k,t,p/) sequences in two phrase positions. The first four spectral coefficients (mean, standard deviation, skewness, kurtosis) were calculated for one window centered on the burst of the onset consonant and two subsequent, non-overlapping windows. All coefficients showed an influence of vowel-to-consonant anticipatory coarticulation. Which onset consonant showed the strongest vowel effect depended on the specific coefficient under consideration. A context-dependent consonant-to-consonant anticipatory effect was observed for onset /p/. Findings demonstrate that spectral coefficients can reveal subtle anticipatory adjustments as early as the burst of stop consonants. Different results for the four coefficients suggest that comprehensive spectral analyses offer advantages over other approaches. Studies using these techniques may expose previously unobserved articulatory adjustments among phonetic contexts or speaker populations.

Keywords

Speech production; Articulation; Anticipatory coarticulation; Acoustic analysis; Spectral coefficients

1. Introduction

Coarticulation was mentioned in the literature as early as 1933 (Menzerath and de Lacerda, 1933) and has received considerable research attention since the late 1960s (Daniloff and Moll, 1968; Öhman, 1966, 1967; Kent and Minifie, 1977). Both empirical and theoretical

work has emphasized anticipatory (rather than perseverative) effects because such anticipatory adjustments could provide important information about the central processes involved in speech planning (Whalen, 1990). In particular, numerous investigations have documented anticipatory vowel-to-vowel coarticulation across an intervening consonant (e.g., Öhman, 1966, 1967; Purcell, 1979; Recasens, 1984b, 1987), the effect of a consonant on a preceding vowel or vice versa (e.g., Farnetani and Recasens, 1993; MacNeilage and DeClerk, 1969; Ostry et al., 1996; Parush et al., 1983), and the effect of a symmetric consonant context on a medial vowel (e.g., House and Fairbanks, 1953; Stevens and House, 1963). Surprisingly, despite demonstrations that a consonant can affect an adjacent consonant (Repp, 1983; Repp and Mann, 1981, 1982) and that the influence of a consonant can extend to the onset of a preceding vowel (Broad and Fertig, 1970; Moll and Daniloff, 1971), only two studies (to the best of our knowledge) have explored consonant-to-consonant anticipatory coarticulation across an intervening vowel. Based on EMG recordings from a single subject producing C_1VC_2 syllables, MacNeilage and DeClerk (1969) failed to observe a transvocalic consonant-to-consonant influence. Only for a syllable's final consonant *voicing* (i.e., a non-articulatory feature) have effects on the acoustic characteristics and perception of the initial consonant been reported (Hawkins and Nguyen, 2000, 2004).

It is important to note, however, that the focus on specific aspects of coarticulation may have been based, at least in part, on the availability of certain acoustic analysis procedures. For example, many investigators have extracted the individual formant frequencies of target acoustic segments to determine the extent to which those formant frequencies are influenced by the segment's phonetic context (e.g., Joos, 1948; Öhman, 1966; Repp, 1986; Repp and Mann, 1982; Sereno et al., 1987; Siren and Wilcox, 1995; Tjaden, 2000). For CV syllables with a stop consonant, other investigators have calculated *locus equations* which reflect the relationship between the frequency of the second formant at the transition onset and the frequency of the same formant in the steady-state portion of the vowel (Sussman et al., 1997, 1992). These approaches have contributed invaluable information regarding specific types of coarticulation in specific syllable shapes, but, for technical reasons, their application is limited to speech segments in which one or more formants can be unambiguously identified.

At the present time, additional analysis techniques are available that may provide more comprehensive information about subtle acoustic characteristics of consonant production (e.g., taking into account the entire spectral distribution) and, importantly, that can be applied to speech segments without identifiable formants such as voiceless segments and the initial burst portion of both voiced and voiceless consonants. Interestingly, several investigators have already used the overall spectral distribution of consonants to examine acoustic invariance across different vowel contexts, and they have demonstrated that those methods can indeed accurately classify the consonants according to place of articulation (Blumstein and Stevens, 1979; Forrest et al., 1988; Kewley-Port, 1983; Stevens and Blumstein, 1978). Thus, those researchers' success in classifying stop consonants shows that the spectral distribution can be used effectively to detect acoustic *similarities* within one single place of articulation (as compared with other places of articulation) despite the different contexts in which the consonant was produced. Those findings raise the question – addressed in the present work – whether acoustic measures based on the overall spectral distribution are also sufficiently sensitive to reveal acoustic *differences* within one single place of articulation when a consonant is produced across different phonetic contexts (i.e., to detect subtle but systematic coarticulation-induced *variability* within the categories that were previously shown to be characterized by relative invariance).

This question remains largely unexplored as very few studies have used the overall spectral distribution at the time of the initial stop release (as opposed to later time points such as during the aspiration noise or formant transition) to investigate vocal tract adjustments as a result of anticipatory coarticulation. Using squared difference FFT spectra, Hertrich and Ackermann (1999) described anticipatory CV as well as perseverative VC coarticulation effects for German speakers with and without cerebellar disorders. A different approach may consist of quantifying multiple aspects of the spectral shape with a small set of statistical indices. Our rationale for this spectral statistics approach is based on work showing that the first four spectral coefficients¹ (i.e., mean, standard deviation, skewness, and kurtosis, which describe the spectrum's central tendency, variation, symmetry, and peakedness, respectively) can be used to reveal acoustic differences among obstruents (Forrest et al., 1988). Initial applications of such spectral coefficient analyses to the study of coarticulation in the burst and early part of the release of the stops /k/ and /t/ (but not /p/) have been published by Tjaden (2003) and Tjaden and Wilding (2005) who calculated only the spectral mean for adults with versus without neurological disease and also by Nittrouer (1995) who calculated the spectral mean, skewness, and kurtosis for normal children versus adults.

There have been no studies, however, that have investigated anticipatory coarticulation in the speech of non-disordered adult speakers by calculating all four spectral coefficients for each of the three places of articulation (palatal/velar, alveolar, bilabial) that are applicable to American English voiceless stop consonants (/k,t,p/). Moreover, none of the previous studies have explored the possibility that spectral coefficients may be able to reveal, for the first time, transvocalic consonant-to-consonant anticipatory coarticulation. Indeed, calculations of all four coefficients for the burst and early part of the release of stops in C₁VC₂ syllables could offer an additional window into previously documented (e.g., V–C₁) as well as previously undocumented (e.g., C₂–C₁) anticipatory coarticulation effects – thereby providing new data with potential theoretical and clinical implications. Consequently, the purpose of the present study was threefold: (a) to examine if, and which, spectral coefficients are sensitive enough to reveal known vowel-related coarticulation effects on each of the three voiceless stop consonants occurring in syllable-initial position, (b) to examine if, and which, spectral coefficients reflect such anticipatory coarticulation as early as the burst of the stop consonant, and (c) to examine if, and which, spectral coefficients can show an effect not only of the subsequent vowel but also of the syllable-final stop consonant. In sum, this study was designed to investigate the presence or absence of anticipatory coarticulation in spectral coefficient measures of the initial stop consonant in C₁VC₂ syllables consisting of various combinations of initial stop consonants (onset), vowels (nucleus), and final stop consonants (coda).

2. Material and methods

2.1. Participants

Participants were twenty adults who were native speakers of American English and who resided in the North East of the United States (Connecticut). The group included 10 female and 10 male subjects. Female subjects ranged in age from 22 to 34 years (mean age 24.3 years). Male subjects ranged in age from 19 to 27 years (mean age 22.7 years). None of the subjects reported a history of speech, hearing, or language disorders. Each subject gave informed consent before participating.

¹Throughout this work we use the term spectral coefficients, rather than the commonly used term spectral moments, because, by definition, the coefficients are formulated from, and thus are different from, the moments (Newell and Hancock, 1984).

2.2. Procedure

Subjects read from a computer monitor individually displayed carrier phrases with embedded target monosyllables. The targets were always preceded by the article “a” (per the experimenter’s instructions pronounced as schwa) and of the form C_1VC_2 in which C_1 was a palatal/velar, alveolar, or bilabial voiceless stop consonant (/k/, /t/, /p/), V was a high front, low front, or mid back vowel (/i/, /æ/, /o/; orthographically represented for the subjects as *ee*, *a*, *oa*, respectively), and C_2 was again a palatal/velar, alveolar, or bilabial voiceless stop consonant. Thus, target syllables represented the following nine C_1C_2 patterns for each of the three intervening vowels V :

kVk	kVt	kVp
tVk	tVt	tVp
pVk	pVt	pVp

Given the inclusion of all systematic combinations, the pronunciation of some syllables corresponded to that of an existing American English word, but most were non-words. To allow an examination of possible sentence position effects, each of the 27 combinations was placed in two different positions within a carrier phrase (“*He says it is a _.*” and “*A _ is what he says.*”). For both positions, subjects were instructed to place sentence stress on the target syllable. Subjects produced four trials of each of these 54 utterances – resulting in a total of 216 trials per subject. The trial order was randomized for each subject individually, and no instructions were given with regard to loudness or rate.²

Using a mouth-to-microphone distance of 15 cm, the acoustic signal transduced with a microphone (SHURE SM58) was amplified (ART DPS DI/O Preamp System) and digitized directly onto CD at 44100 Hz (JVC CD/CDR Recorder XL-R5000BK). The CD audio files were software-converted to “wav” format and digitally down-sampled to 22050 Hz. With only seven trials unavailable for analysis (one subject failed to produce two of the trials, five other subjects each failed to produce one trial), data were extracted from a large database of 4313 trials.

2.3. Data extraction

All data were obtained through spectral analysis of the syllable-initial C_1 consonants. Praat software (<http://www.fon.hum.uva.nl/praat/>) was used to zoom in on the spectrogram of each target syllable and to place a marker at the onset of each C_1 consonant burst. A custom script was used to extract the acoustic signal from –10 to +30 ms relative to each marker. Using custom MATLAB routines, running spectral analyses were then applied to the extracted signals. As illustrated in Fig. 1, the first four coefficients (Newell and Hancock, 1984) were computed from the linear frequency scale spectrum for each of three successive 11.61 ms windows (i.e., each window consisted of 256 samples) with the first window centered on the onset of the burst of the C_1 consonant. Specifically, after pre-emphasizing the digital signal, a 11.61 ms Hamming window was applied, a 256-point Fast Fourier

²Articulation rate for the produced utterances was measured during the data analysis stage, and the mean rate across all subjects was determined to be 5.27 ($SD = .83$) syllables per second, with no statistically significant difference between female and male subjects (5.09 and 5.46 syllables per second, respectively). These articulation rate measures were automatically extracted using custom-written Matlab code. For each subject, the procedure involved measuring the total duration of all intervals that did not contain breaks longer than 300 ms (breaks were defined as periods where the acoustic energy dropped below a threshold value), and then dividing this duration by the number of spoken syllables. The 300 ms criterion was based on an established pause duration of 250 ms (Goldman-Eisler, 1961; Max and Caruso, 1998) and an additional 50 ms to account for the utterance-final and utterance-initial consonants whose low energy would fall below the specified energy threshold (Rabiner and Sambur, 1975).

Transform (FFT) was computed, and the power spectrum across discrete frequency samples was obtained.

The power spectrum was normalized as:

$$p(f_k) = \frac{P(f_k)}{\sum_{k=0}^{N/2} P(f_k)} \quad (1)$$

where $P(f_k)$ is the power spectrum, $f_k = 2f_{Nq}k/N$, $k = 0, 1, \dots, N/2$, and $N = 256$. f_{Nq} indicates the Nyquist frequency (11025 Hz), and, thus, the upper limit of the frequency range that was used to compute spectral coefficients.

Four coefficients of the normalized power spectrum were computed as:

$$\text{mean, } \mu = \sum_{k=0}^{N/2} p(f_k) f_k \quad (2)$$

$$\text{standard deviation } \sigma = \sqrt{\sum_{k=0}^{N/2} (f_k - \mu)^2 p(f_k)} \quad (3)$$

$$\text{skewness} = \sum_{k=0}^{N/2} \left(\frac{f_k - \mu}{\sigma} \right)^3 p(f_k) \quad (4)$$

$$\text{kurtosis} = -3 + \sum_{k=0}^{N/2} \left(\frac{f_k - \mu}{\sigma} \right)^4 p(f_k) \quad (5)$$

Skewness is positive when there is a relatively greater amount of acoustic energy in the low-frequency range and a relatively smaller amount of acoustic energy in the high-frequency range such that the spectrum is asymmetric with its tail in the high-frequency range. Skewness is negative when there is more energy in the high-frequency range and the tail is in the low frequency range. Skewness is zero for a symmetric spectrum. For kurtosis, values are positive when the overall shape of the spectrum is more peaked than the normal distribution, negative when it is less peaked, and zero when it is normally distributed.

2.4. Statistical analyses

Prior to statistical analysis, each subject's spectral coefficient measures were averaged across the four repetitions of each unique utterance. Then, for each syllable onset separately, an analysis of variance (ANOVA) was completed with nucleus (3 levels), coda (3 levels), carrier phrase position (2 levels), and window (3 levels) as within-subjects variables and gender as a between-subjects variable. Given the exploratory nature of the study, the expected subtlety of some potential effects (e.g., articulatory effects of the coda on the onset have not been detected to date), and the need to balance Types II and I errors in this

situation, an uncorrected level of .01 was selected (Kirk, 1995; Skipper et al., 1972). Results for all repeated measures variables were obtained with Huynh-Feldt epsilon-corrected degrees of freedom to correct for any violations of the sphericity assumption (Max and Onghena, 1999). Homogeneity of variance between the gender groups was tested for each combination of the four repeated measures variables. Using Levene's test, homogeneity was confirmed for approximately 92% of those combinations. Partial eta-squared (hereafter η^2_p) was computed as a measure of effect size (strength of association) for each main effect and interaction in the ANOVAs.

3. Results

3.1. Onset /k/

All descriptive data for the spectral coefficients calculated for onset /k/ are provided in the Supplementary data accompanying this article. Statistically significant inferential results for onset /k/ are included here in Table 1.

For each coefficient, statistically significant main effects were found for the variables window and nucleus. The main effect of nucleus on the first window of onset /k/ is illustrated in Fig. 2. Although only three dimensions can be shown in these graphs (we selected the three spectral coefficients with the largest effect sizes for the nucleus factor), /k/ onset consonants produced in different nucleus contexts (represented with light gray solid lines, dark gray dashed lines, and black dotted lines) are clearly separated in this multidimensional space. In contrast, these /k/ onset consonants are not consistently separated in the same space when produced with a given nucleus in different carrier phrase positions (round vs. square markers) or with different coda contexts (top vs. middle vs. bottom panel). To illustrate the strong nucleus effect also at the individual subject level, Fig. 3 shows each individual speaker's first window data for onset /k/ averaged across the three codas and two phrase positions. Each panel represents a combination of two spectral coefficients, and illustrates that the different nuclei caused a shift in the spectral characteristics of the onset consonant.

The only interaction that reached statistical significance for onset /k/ was between window and nucleus, and this effect was significant for both the spectral mean and kurtosis. As shown in the left column of Fig. 4, the size of the above described nucleus effect on this dependent variable gradually increased across the three windows (i.e., as the analysis window moved closer to the onset of the nucleus).

3.2. Onset /t/

Detailed descriptive data for the spectral coefficients computed for onset /t/ are included in the Supplementary data. Table 2 summarizes the statistically significant inferential results. Main effects were statistically significant for the factors window and nucleus (in each case for two out of four dependent variables), but the interaction between window and nucleus was statistically significant for all four dependent variables. The overall nucleus effect on the first window of onset /t/ is shown in Fig. 5 which shows results generally consistent with those reported above for onset /k/, albeit with a less distinct separation of productions paired with the front vowels /i/ and /æ/. Individual subject data for the first window in onset /t/ are shown in Fig. 6 which again reveals clear shifts in spectral characteristics of the onset consonant when produced in different nuclei contexts.

For all four spectral coefficients, the interaction between window and nucleus is graphically illustrated in the middle column of Fig. 4. The size of the differences among /t/ onsets produced with each nucleus varied across the three windows, although these differences did not always progressively increase as was the case for /k/ onsets.

The coda factor contributed to a statistically significant three-way interaction with window and position. Specifically, in the early utterance position skewness always first decreased (from windows 1 to 2) before it increased to a maximum value (from windows 2 to 3) whereas in the late utterance position skewness always increased gradually (from windows 1 to 3), and in both positions the difference among the three windows was larger for coda /p/ than for codas /k/ and /t/.

3.3. Onset /p/

Descriptive data for the spectral coefficients obtained for onset /p/ are available in the Supplementary data accompanying this article, and all statistically significant inferential results are listed in Table 3. Overall, /p/ productions yielded a greater number of statistically significant results than did either /k/ or /t/. We focus here only on those results that inform specifically about aspects of anticipatory coarticulation (i.e., main effects or interactions involving the factors window, nucleus, or coda), but all results are included in the table as they may be informative for future studies (for example, the strong position effect for this consonant).

The main effects of window and nucleus were statistically significant for all four spectral coefficients, but so was the interaction between these two independent variables. The main effect of nucleus on the first window of onset /p/ is illustrated in Fig. 7. Although statistically significant and associated with large effect sizes for each of the three dependent variables used to create the plots, the separation of the group average data for different nucleus contexts is visually less obvious as compared with those shown above for onset /k/ and onset /t/. Individual subject data for the /p/ onset are presented in Fig. 8.

Similar to the situation for onset /k/, the interaction between window and nucleus for all coefficients reflected, generally, a progressively increasing effect of nucleus from the first to the third analysis window as can be seen in the right-side column of Fig. 4. Other interactions were statistically significant for one or two dependent variables. For example, the interaction between window and gender for the first coefficient was caused by male and female subjects showing very similar spectral means in the first two windows but more differentiated spectral means in the third window when the nucleus influence became more substantial.

Of particular interest for the present purpose are interactions involving the coda factor. First, there was a statistically significant ($p < .01$) coda $\times$ nucleus interaction for the third spectral coefficient (the same interaction resulted in $p = .012$ for the first spectral coefficient) as shown in Fig. 9. Given that this interaction was not further modified by window, Fig. 9 is based on data from only the first window in order to illustrate coarticulation that occurred as early as the burst of the onset consonant: the third spectral coefficient of these productions of /p/ showed a coda effect that varied in direction across the different nuclei. Specifically, additional testing for each nucleus separately revealed a strong trend toward a coda effect on the third spectral coefficient of /p/ for productions paired with /æ/ ($F(1.893, 35.967) = 4.674$, $p = .017$, $\eta^2_p = .197$) but not /i/ or /o/ ($p > .1$).

The coda factor also contributed to a second statistically significant interaction. Coda, window, and gender interacted in a three-way effect on the fourth spectral coefficient (and showed a trend toward a three-way effect for the second spectral coefficient, with $p = .02$): descriptively, the strongest trends toward a coda effect on /p/ kurtosis occurred in the third window for the female speakers but in the second window for the male speakers.

4. Discussion

The aims of this study were to determine whether spectral coefficients can serve as acoustic indices of the anticipatory influence of a nucleus and coda stop consonant on the burst and early aspiration of the onset stop consonant in CVC syllables and, if so, to document which spectral coefficients can reliably quantify the acoustic changes associated with such coarticulation effects. First, one primary finding of this work is that the spectral coefficients did indeed reveal several coarticulation effects, including some that had already been documented in previous reports based on different methodologies and others that have not been shown in studies to date. Second, the spectral coefficient measures allowed these anticipatory coarticulation effects to be shown as early as the syllable-onset release burst and also during the early part of aspiration noise. Third, specific coarticulation effects of interest included a relatively consistent main effect of nucleus on most spectral coefficients (all four for /k/ and /p/, two for /t/) and a relatively consistent interaction between nucleus and analysis window on most spectral coefficients (all four for /t/ and /p/, two for /k/), an interaction between nucleus and coda on the third spectral coefficient of onset consonant /p/, and some higher-order interactions of coda and analysis window with position of the target within the carrier phrase (third coefficient of /t/) or with gender (fourth coefficient of /p/). Fourth, some of these effects, as well as the relative strength of coarticulation observed in a given consonant (as quantified by the effect size measures), depended on which of the four coefficients was considered. Below, these primary findings are discussed in more detail within the overall context of vowel-to-consonant anticipatory coarticulation (nucleus effects) and consonant-to-consonant anticipatory coarticulation (coda effects).

4.1. Vowel-to-consonant anticipatory coarticulation (nucleus effects)

The finding of a strong and consistent nucleus effect is in keeping with a large body of the literature on vowel-to-consonant anticipatory coarticulation in CV(C) syllables. Using perceptual, acoustic, kinematic, or electromyographic data, numerous investigators have demonstrated such an anticipatory influence of a vowel on a preceding consonant (e.g., Löfqvist and Gracco, 1994; MacNeilage and DeClerk, 1969; Ostry et al., 1996; Parush et al., 1983; Repp, 1986; Sereno et al., 1987; Stevens and Blumstein, 1978; Sussman et al., 1997; Winitz et al., 1972). The present replication of such previous findings validates the use of spectral coefficients as dependent variables that can detect coarticulatory influences.

Here, the use of spectral analyses applied to the aperiodic noise of voiceless stop consonants allowed anticipatory coarticulation to be revealed as early as the burst associated with the release of the articulatory obstruction (i.e., the first analysis window). The importance of this finding lies in the fact that it demonstrates advantages of spectral coefficients analyses over formant-based methods given that the latter (a) are typically used to measure a vowel's influence on a stop consonant at the onset of the quasi-periodic signal associated with the vowel itself (Repp, 1986; Sussman et al., 1997, 1992), and (b) do not provide information about the overall spectral distribution. These methodological advantages may make it possible for future studies based on spectral coefficients to expose previously undetected differences in stop consonant articulation across phonetic contexts or between separate groups of speakers (e.g., individuals with normal versus disordered speech, young versus elderly speakers).

To maximize the value of such future work, it will be important to further explore the relationship between the spectral coefficients (individually and in combination) and specific vocal tract configurations. Although we recognize that the various indices describing a distribution are inter-dependent and that each spectral coefficient is not necessarily individually related to a specific vocal tract parameter, there is currently a paucity of information about the relationship between the spectral standard deviation, skewness, and

kurtosis and vocal tract configuration. The spectral mean, on the other hand, has been documented to vary with the size of the cavity in front of an articulatory constriction: the longer the anterior cavity, the lower the spectral mean (Nitttrouer, 1995; Stevens, 1998). At least two findings of the present work are in keeping with this previously demonstrated interpretation of the spectral mean. First, we found descriptively higher spectral means for /t/ than for /k/ in every nucleus and coda context and in both phrase positions (see Supplementary data for details), and this is in agreement with the prediction that the shorter anterior cavity for /t/ than for /k/ should cause an upward shift of the mean spectral frequency (note that the relation between mean spectral frequency and place of articulation cannot be applied to nonlingual consonants such as /p/, see Gordon et al., 2002). Second, the spectral means for /t/ and /k/ produced before /o/ were lower than before /æ/ and /i/ (see Supplementary data). Although an influence of lip rounding cannot be ruled out, this finding is consistent with the prediction that the coarticulatory influence of the back vowel /o/ would result in a more posterior lingual obstruction for the consonant (longer anterior cavity), whereas the front vowels /æ/ and /i/ would result in a more anterior lingual obstruction for the consonant (shorter anterior cavity).

Our data for spectral standard deviation, skewness, and kurtosis suggest that these coefficients may be affected, in part, by the amount of lip rounding versus spreading. That is, as the lip configuration varied from rounded (vowel /o/) to spread (vowel /i/), the spectral skewness and kurtosis generally increased, whereas the spectral standard deviation generally decreased. For example, as the lip configuration varied from rounded to spread, the average skewness for /k/ (preceding /o/ 0.75, /æ/ 1.18, /i/ 1.30) and average kurtosis for /k/ (preceding /o/ 0.54, /æ/ 1.85, /i/ 2.64) increased while the average standard deviation for /k/ decreased (preceding /o/ 2.51, /æ/ 1.95, /i/ 1.76). The same trends were also observed for the skewness, kurtosis, and standard deviation for /p/ (see bottom row of the tables provided as Supplementary data). However, this apparent relation to lip rounding did not apply to all vocal tract configurations given that the effect of rounding versus spreading on the standard deviation of /t/ (preceding /o/ 1.54, /æ/ 1.55, /i/ 1.48), skewness of /t/ (preceding /o/ 1.12, /æ/ 0.63, /i/ 0.83), and kurtosis of /t/ (preceding /o/ 3.04, /æ/ 2.11, /i/ 2.59) differed from that observed for /k/ and /p/.

It is also of interest that the degree of vowel-to-consonant anticipatory coarticulation (as quantified by the effect sizes of the nucleus main effects) for the three different onset consonants depends on the specific spectral coefficient under consideration. For instance, when the spectral mean is considered, /t/ showed more coarticulation (or, in the terminology of Recasens (1985), less coarticulation *resistance*) than /k/ and /p/; when the spectral standard deviation is considered, /k/ and /p/ showed similarly high effect sizes in the absence of an effect for /t/; when the spectral skewness and kurtosis are considered, /p/ showed more coarticulation than /t/ and /k/. Based on electropalato-graphic (EPG), acoustic, and kinematic data, others have rank ordered consonants with different places or manner of articulation in terms of their relative resistance to coarticulation (e.g., Fowler and Brancazio, 2000; Recasens, 1984a,b, 1985; Recasens et al., 1997; Sussman, 1994; Sussman et al., 1991). However, the effect sizes for the present acoustic results – expressing the proportion of the variance in spectral coefficient measures of onset stop consonants that is accounted for by the subsequent nuclei – suggest that there may be no single place of articulation that is most resistant to the influence of surrounding articulatory gestures: instead, rank-ordering consonants with different place of articulation in terms of coarticulation resistance may yield different results depending on the analysis technique (e.g., acoustic, EPG, kinematic), the specific dependent variable that is used (e.g., here four different spectral coefficients), the analyzed portion of the consonant (e.g., burst, aspiration noise, formant transition onset), and possibly the language.

4.2. Consonant-to-consonant anticipatory coarticulation (coda effects)

Unlike the frequent findings of nucleus effects on a preceding consonant, reports of coda effects on the onset consonant are extremely scarce. Although Hawkins and Nguyen (2000,2004) found an effect of coda voicing on both the acoustics and the perception of onset /l/, there appear to be no peer-reviewed publications demonstrating anticipatory adjustments based on *articulatory* aspects of a coda consonant. Indeed, in the best known study that focused in part on coda-to-onset coarticulation, MacNeilage and DeClerk (1969) failed to find such effects in EMG data that were collected from a single subject. Therefore, it may be theoretically important that the present work based on a relatively large group of subjects revealed some effects, albeit limited and highly context-dependent, of the coda place of articulation on the spectral characteristics of some onset consonants: the spectral skewness and kurtosis of /p/ exhibited statistically significant coda by nucleus and coda by window by gender interactions, respectively, and the spectral skewness of /t/ exhibited a statistically significant coda by window by position interaction.

Any interpretations of the mechanisms underlying such possible coda effects, and the modulating influence of nucleus and gender, will have to await further replication of these preliminary findings and, if confirmed, further exploration with kinematic analyses. Nevertheless, the finding that a coda by nucleus interaction was found only for onset /p/ – and, in fact, also the findings that a larger number of statistically significant results and large effect sizes for the nucleus influence were found for onset /p/ – may have occurred for at least two possible reasons. One possibility is based on well-established findings that the voice onset time (VOT) is typically shorter for /p/ than for /k/ or /t/ (Cho and Ladefoged, 1999; Lisker and Abramson, 1964). Therefore, the bilabial consonant's shorter interval between the onset burst and the nucleus could result in stronger coarticulation than that for the palatal/velar and alveolar onset consonants. A second possibility is that the bilabial place of articulation imposes fewer constraints on – and, thus, allow an earlier onset of – lingual gestures toward the articulatory targets for the nucleus and two of the three coda consonants. These hypotheses regarding earlier movements toward various nucleus and coda targets for onset /p/ can be empirically tested in future kinematic experiments.

4.3. Conclusion

In summary, vowel-to-consonant and transvocalic consonant-to-consonant anticipatory coarticulation in CVC monosyllables were examined by means of detailed spectral coefficient analyses. Both these types of coarticulatory effects could be observed as early as the release burst of voiceless stop consonants in syllable-initial position. The appropriateness of spectral coefficient analyses was confirmed by the large effect sizes obtained in replicating, for all onset consonants, well-known vowel-to-consonant anticipatory coarticulation effects. The relative strength of the nucleus effect differed across the onset consonants depending on the spectral coefficient under consideration. Thus, the vowel-to-consonant results suggest that a given consonant's degree of coarticulation depends on the specific dependent variables used to quantify coarticulation. This finding emphasizes that, even though the four statistical coefficients are interdependent to some extent, they are also sufficiently independent so that different coefficients can capture the coarticulatory influence of different vocal tract configurations. In other words, obtaining different results for different coefficients confirmed the expected benefits of including a larger set of spectral coefficients as compared with previous coarticulation studies. Moreover, the analysis methods employed were also shown to be sufficiently sensitive to reveal a consonant-to-consonant coarticulation effect in one phonetic context – an effect not previously shown with either acoustic or physiological techniques. Taken together, the findings reported here suggest that future applications of spectral coefficient analyses – and possibly other related techniques based on whole spectrum analysis (e.g., based on the

multidimensional 1/3 octave filter approach described by Plomp et al. (1967) – may provide investigators with more refined tools to address both theoretically- and clinically-motivated studies of coarticulatory effects in the production of stop consonants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by Grant R01DC007603 from the National Institute on Deafness and Other Communication Disorders (L. Max PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.specom.2011.02.003.

References

- Blumstein SE, Stevens KN. Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* 1979; 66:1001–1017. [PubMed: 512211]
- Broad DJ, Fertig RH. Formant-frequency trajectories in selected CVC syllable nuclei. *J. Acoust. Soc. Am.* 1970; 47:1572–1582. [PubMed: 5426309]
- Cho T, Ladefoged P. Variation and universals in VOT: evidence from 18 languages. *J. Phonetics.* 1999; 27:207–229.
- Daniloff R, Moll K. Coarticulation of lip rounding. *J. Speech Lang. Hear. Res.* 1968; 11:707–721.
- Farnetani E, Recasens D. Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Lang. Speech.* 1993; 36:279–302. [PubMed: 8277811]
- Forrest K, Weismer G, Milenkovic P, Dougall RN. Statistical analysis of word-initial voiceless obstruents: preliminary data. *J. Acoust. Soc. Am.* 1988; 84:115–123. [PubMed: 3411039]
- Fowler CA, Brancazio L. Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Lang. Speech.* 2000; 43:1–41.
- Goldman-Eisler F. Continuity of speech utterance, its determinants and its significance. *Lang. Speech.* 1961; 4:220–231.
- Gordon M, Barthmaier P, Sands K. A cross-linguistic acoustic study of voiceless fricatives. *J. Int. Phonetic Assoc.* 2002; 32:141–174.
- Hawkins, S.; Nguyen, N. In: Cutler, A.; McQueen, JM.; Zondervan, R., editors. Predicting syllable-coda voicing from the acoustic properties of syllable onsets; SWAP: The ISCA Conference on Spoken Word, Access Processes; ISCA Archive, Nijmegen, The Netherlands. 2000; p. 167-170.
- Hawkins S, Nguyen N. Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *J. Phonetics.* 2004; 32:199–231.
- Hertrich I, Ackermann H. Temporal and spectral aspects of coarticulation in ataxic dysarthria: an acoustic analysis. *J. Speech Lang Hear. Res.* 1999; 42:367–381. [PubMed: 10229453]
- House AS, Fairbanks G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 1953; 25:105–113.
- Joos M. Acoustic Phonetics. *Language.* 1948; 24:1–136.
- Kent RD, Minifie FD. Coarticulation in recent speech production models. *J. Phonetics.* 1977; 5:115–133.
- Kewley-Port D. Time-varying features as correlates of place of articulation in stop consonants. *J. Acoust. Soc. Am.* 1983; 73:322–335. [PubMed: 6826902]

- Kirk, RE. *Experimental Design: Procedures for the Behavioral Sciences*. Brooks/Cole; Pacific Grove, CA: 1995.
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: acoustic measurements. *Word*. 1964; 20:384–422.
- Löfqvist A, Gracco VL. Tongue body kinematics in velar stop production: influences of consonant voicing and vowel context. *Phonetica*. 1994; 51:52–67. [PubMed: 8052675]
- MacNeilage PF, DeClerk JL. On the motor control of coarticulation in CVC monosyllable. *J. Acoust. Soc. Am.* 1969; 45:1217–1233. [PubMed: 5803161]
- Max L, Caruso AJ. Adaptation of stuttering frequency during repeated readings: associated changes in acoustic parameters of perceptual fluent speech. *J. Speech Lang. Hear. Res.* 1998; 41:1265–1281. [PubMed: 9859883]
- Max L, Onghena P. Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *J. Speech Lang. Hear. Res.* 1999; 42:261–270. [PubMed: 10229445]
- Menzerath, P.; de Lacerda, A. *Koartikulation, Steuerung und Lautabgrenzung*. Ferd, Dummlers Verlag; Berlin/Bonn: 1933.
- Moll KL, Daniloff RG. Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.* 1971; 50:678–684. [PubMed: 5096503]
- Newell KM, Hancock PA. Forgotten moments: a note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions. *J. Motor Behav.* 1984; 16:320–335.
- Nittrouer S. Children learn separate aspects of speech production at different rates: evidence from spectral moments. *J. Acoust. Soc. Am.* 1995; 97:520–530. [PubMed: 7860830]
- Öhman SEG. Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.* 1966; 39:151–168. [PubMed: 5904529]
- Öhman SEG. Numerical model of coarticulation. *J. Acoust. Soc. Am.* 1967; 41:310–320. [PubMed: 6040806]
- Ostry DJ, Gribble PL, Gracco VL. Coarticulation of jaw movements in speech production: is context sensitivity in speech kinematics centrally planned? *J. Neurosci.* 1996; 16:1570–1579. [PubMed: 8778306]
- Parush A, Ostry DJ, Munhall KG. A kinematic study of lingual coarticulation in VCV sequences. *J. Acoust. Soc. Am.* 1983; 74:1115–1125. [PubMed: 6643833]
- Plomp R, Pols LCW, van de Geer JP. Dimensional analysis of vowel spectra. *J. Acoust. Soc. Am.* 1967; 41:707–712.
- Purcell ET. Formant frequency patterns in Russian VCV utterances. *J. Acoust. Soc. Am.* 1979; 66:1691–1702. [PubMed: 521553]
- Rabiner LR, Sambur MR. An algorithm for determining the endpoints of isolated utterances. *Bell Syst. Techn. J.* 1975; 54:297–315.
- Recasens D. V-to-C coarticulation in Catalan VCV sequences: an articulatory and acoustical study. *J. Phonetics*. 1984a; 12:61–73.
- Recasens D. Vowel-to-vowel coarticulation in Catalan VCV sequences. *J. Acoust. Soc. Am.* 1984b; 76:1624–1635. [PubMed: 6520300]
- Recasens D. Coarticulatory patterns and degrees of coarticulation resistance in Catalan CV sequences. *Lang. Speech*. 1985; 28:97–114. [PubMed: 4087987]
- Recasens D. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *J. Phonetics*. 1987; 15:299–312.
- Recasens D, Pallares MD, Fontdevila J. A model of lingual coarticulation based on articulatory constraints. *J. Acoust. Soc. Am.* 1997; 102:544–561.
- Repp BH. Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *J. Acoust. Soc. Am.* 1983; 74:420–427. [PubMed: 6619419]
- Repp BH. Some observations on the development of anticipatory coarticulation. *J. Acoust. Soc. Am.* 1986; 79:1616–1619. [PubMed: 3711460]
- Repp BH, Mann VA. Perceptual assessment of fricative-stop coarticulation. *J. Acoust. Soc. Am.* 1981; 69:1154–1163. [PubMed: 7229203]

- Repp BH, Mann VA. Fricative-stop coarticulation: acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 1982; 71:1562–1567. [PubMed: 7108031]
- Sereno JA, Baum SR, Marean GC, Lieberman P. Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *J. Acoust. Soc. Am.* 1987; 81:512–519. [PubMed: 3558969]
- Siren KA, Wilcox KA. Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *J. Speech Lang. Hear. Res.* 1995; 38:351–359.
- Skipper, JK.; Guenther, AL.; Nass, G. The sacredness of .05: a note concerning the use of statistical levels of significance in social science. In: Kirk, RE., editor. *Statistical Issues: A Reader for the Behavioral Sciences*. Brooks/Cole; Monterey, CA: 1972. p. 141-145.
- Stevens, KN. *Acoustic Phonetics*. The MIT Press; Cambridge, MA: 1998.
- Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 1978; 64:1358–1368. [PubMed: 744836]
- Stevens KN, House AS. Perturbations of vowel articulations by consonantal context: an acoustical study. *J. Speech Lang. Hear. Res.* 1963; 6:111–128.
- Sussman HM. The phonological reality of locus equations across manner class distinctions: preliminary observations. *Phonetica*. 1994; 51:119–131.
- Sussman HM, Bessell N, Dalston E, Majors T. An investigation of stop place of articulation as a function of syllable position: a locus equation perspective. *J. Acoust. Soc. Am.* 1997; 101:2826–2838. [PubMed: 9165737]
- Sussman HM, Hoemeke K, McCaffrey HA. Locus equations as an index of coarticulation for place of articulation distinctions in children. *J. Speech Lang. Hear. Res.* 1992; 35:769–781.
- Sussman HM, McCaffrey HA, Matthews SA. An investigation of locus equations as a source of relational invariance for stop place categorization. *J. Acoust. Soc. Am.* 1991; 90:1309–1325.
- Tjaden K. An acoustic study of coarticulation in dysarthric speakers with Parkinson's disease. *J. Speech Lang. Hear. Res.* 2000; 43:1466–1480. [PubMed: 11193966]
- Tjaden K. Anticipatory coarticulation in multiple sclerosis and Parkinson's disease. *J. Speech Lang. Hear. Res.* 2003; 46:990–1008. [PubMed: 12959475]
- Tjaden K, Wilding GE. Effect of rate reduction and increased loudness on acoustic measures of anticipatory coarticulation in multiple sclerosis and Parkinson's disease. *J. Speech Lang. Hear. Res.* 2005; 48:261–277. [PubMed: 15989391]
- Whalen DH. Coarticulation is largely planned. *J. Phonetics*. 1990; 18:3–35.
- Winitz H, Scheib ME, Reeds JA. Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *J. Acoust. Soc. Am.* 1972; 51:1309–1317. [PubMed: 5032948]

Fig. 1.

Top: Waveform of a production of *koat* with the 34.83 ms section for analysis marked. Middle: Zoomed display of the analyzed segment. Three successive 11.61 ms Hamming windows (shown with arbitrary amplitude) mark the portions for spectral analysis (labeled *w1*, *w2*, and *w3*). Bottom: FFT spectra of *w1*, *w2*, and *w3*. Line types and colors (solid black, dashed black, solid grey) correspond to those used to indicate time windows in the second panel.

Fig. 2.

Group means for the second (*SD* for standard deviation), third (*Skew* for skewness), and fourth (*Kurt* for kurtosis) spectral coefficients in the first analysis window for productions of onset /k/ when the target CVC occurred as the second syllable (*P2* for Position 2) or sixth syllable (*P6* for Position 6) in the carrier phrase. Data are shown separately for the coda contexts /k/ (top), /t/ (middle), and /p/ (bottom).

Fig. 3.

Individual subject data (averaged across coda contexts and phrase positions) based on all possible combinations of two spectral coefficients (M for mean, SD for standard deviation, $Skew$ for skewness, and $Kurt$ for Kurtosis) in the first window for onset /k/. Data for each nucleus context are plotted with different markers and ellipses. Ellipses enclose 68% of the subjects with the ratio of the ellipses' major and minor axes corresponding to the ratio of the standard deviations of the subjects' coordinates relative to these axes.

Fig. 4.

Nucleus by window interaction effects on four spectral coefficients (*M* for mean, *SD* for standard deviation, *Skew* for skewness, and *Kurt* for Kurtosis) for onsets /k/, /t/, and /p/. Data for each nucleus context are plotted with different markers. The three windowed sections for which data are extracted are labeled *w1*, *w2*, and *w3*.

Fig. 5.

Group means for the second (*SD* for standard deviation), third (*Skew* for skewness), and fourth (*Kurt* for kurtosis) spectral coefficients in the first analysis window for productions of onset /t/ when the target CVC occurred as the second syllable (*P2* for Position 2) or sixth syllable (*P6* for Position 6) in the carrier phrase. Data are shown separately for the coda contexts /k/ (top), /t/ (middle), and /p/ (bottom).

Fig. 6. Individual subject data (averaged across coda contexts and phrase positions) based on all possible combinations of two spectral coefficients (M for mean, SD for standard deviation, $Skew$ for skewness, and $Kurt$ for Kurtosis) in the first window for onset /t/. Data for each nucleus context are plotted with different markers and ellipses. Ellipses enclose 68% of subjects with the ratio of the axes corresponding to that of the standard deviations of subjects' coordinates relative to these axes.

Fig. 7.

Group means for the second (*SD* for standard deviation), third (*Skew* for skewness), and fourth (*Kurt* for kurtosis) spectral coefficients in the first analysis window for productions of onset /p/ when the target CVC occurred as the second syllable (*P2* for Position 2) or the sixth syllable (*P6* for Position 6) in the carrier phrase. Data are shown separately for the coda contexts /k/ (top), /t/ (middle), and /p/ (bottom).

Fig. 8.

Individual subject data (averaged across coda contexts and phrase positions) based on all combinations of two spectral coefficients (M = mean, SD = standard deviation, $Skew$ = skewness, $Kurt$ = Kurtosis) in the first window for onset /p/. Data for each nucleus are plotted with different symbols and ellipses. Ellipses enclose 68% of subjects with the ratio of the axes corresponding to that of the standard deviations of subjects' coordinates relative these axes.

Fig. 9.

Group data for the third spectral coefficient (skewness) in the first analysis window for onset /p/ in different coda contexts. Data are shown separately for the nucleus contexts /æ / (top), /i/ (middle), and /o/ (bottom). Bar heights indicate group means, error bars indicate standard error of the mean.

Table 1

F values (*F*), *p* values (*p*), and effect sizes (η^2P) for onset consonant /k/ (only results with $p < .01$ are included)

Factor	Coefficient	<i>F</i> , <i>p</i>	η^2P
Window	1	$F(1.820, 32.757) = 41.713, p < .001$	.699
	2	$F(1.723, 31.009) = 14.902, p < .001$	.453
	3	$F(1.500, 27.004) = 11.944, p = .001$	.399
	4	$F(1.371, 24.682) = 19.386, p < .001$	.519
Nucleus	1	$F(1.429, 25.727) = 11.271, p = .001$	.385
	2	$F(2.000, 36.000) = 100.096, p < .001$	.848
	3	$F(1.405, 25.292) = 21.366, p < .001$	.543
	4	$F(1.700, 30.596) = 22.788, p < .001$	.559
Window	1	$F(3.440, 61.911) = 4.072, p = .008$	.184
× nucleus	4	$F(3.067, 55.210) = 7.063, p < .001$	.282

Table 2

F values (*F*), *p* values (*p*), and effect sizes (η^2P) for onset consonant /*v*/ (only results with $p < .01$ are included)

Factor	Coefficient	<i>F</i> , <i>p</i>	η^2P
Window	1	$F(1.896, 34.127) = 189.431, p < .001$	.913
	4	$F(1.561, 28.102) = 18.016, p < .001$	.500
Nucleus	1	$F(2.000, 36.000) = 61.318, p < .001$	.773
	3	$F(1.557, 28.019) = 28.499, p < .001$	.613
Window $\times$ nucleus	1	$F(4.000, 72.000) = 10.777, p < .001$	.375
	2	$F(2.756, 49.617) = 40.011, p < .001$	.690
	3	$F(2.234, 40.211) = 25.619, p < .001$	.587
Window $\times$ gender	4	$F(1.950, 35.105) = 36.351, p < .001$	.669
	2	$F(1.900, 34.205) = 7.490, p = .002$	.294
Window $\times$ position	3	$F(1.846, 33.230) = 6.115, p = .007$	.254
Coda $\times$ window $\times$ position	3	$F(3.795, 68.317) = 4.433, p = .004$	.198

Table 3

F values (*F*), *p* values (*p*), and effect sizes (η^2P) for onset consonant /p/ (only results with $p < .01$ are included)

Factor	Coefficient	F, p	η^2P
Window	1	$F(2,000, 36,000) = 262.217, p < .001$	.936
	2	$F(1,659, 29,853) = 103.400, p < .001$	.852
	3	$F(2,000, 36,000) = 141.393, p < .001$	.887
	4	$F(1,865, 33,572) = 104.200, p < .001$	.853
Nucleus	1	$F(1,967, 35,412) = 9.307, p = .001$	.341
	2	$F(2,000, 36,000) = 77.517, p < .001$	.812
	3	$F(2,000, 36,000) = 67.886, p < .001$	.790
	4	$F(2,000, 36,000) = 54.176, p < .001$	.751
Position	1	$F(1,000, 18,000) = 39.785, p < .001$	.688
	2	$F(1,000, 18,000) = 11.702, p = .003$	.394
	3	$F(1,000, 18,000) = 21.421, p < .001$	.543
	4	$F(1,000, 18,000) = 21.398, p < .001$	.543
Gender	2	$F(1,000, 18,000) = 9.525, p = .006$	.346
Window $\times$ gender	1	$F(2,000, 36,000) = 5.432, p = .009$	.232
Window $\times$ nucleus	1	$F(3,439, 61,900) = 11.529, p < .001$	.390
	2	$F(4,000, 72,000) = 38.418, p < .001$	.681
	3	$F(3,849, 69,280) = 26.237, p < .001$	.593
	4	$F(3,885, 69,936) = 30.858, p < .001$	.632
Window $\times$ position	4	$F(1,886, 33,950) = 5.601, p = .009$	.237
Coda $\times$ nucleus	3	$F(4,000, 72,000) = 3.713, p = .008$	.171
Coda $\times$ window $\times$ gender	4	$F(4,000, 72,000) = 6.385, p < .001$	.262
Nucleus $\times$ position $\times$ gender	1	$F(2,000, 36,000) = 5.884, p = .006$	.246
	3	$F(2,000, 36,000) = 5.551, p = .008$	.236