ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb56. Examining the extent of anticipatory coronal coarticulation: A long-term average spectrum analysis
Alexei Kochetov and Chris Neufeld*

*Corresponding author's address: Speech Language Pathology, University of Toronto, Toronto, M5G 1V7, Ontario, Canada, christopher.neufeld@utoronto.ca

Phonetic studies of English liquids /r/ and /l/ have shown these consonants can exert strong coarticulatory effects on both adjacent and non-adjacent vowels. The current study investigated local and long-range effects of coronals /l/, /r/, and /d/ in Canadian English. Fourteen speakers were recorded reading the sentences 'We thought it might be a ram/lamb/dam/ham'. Formants F1-F3 and long-term average spectra (LTAS) of 5 vowels preceding the target consonants were calculated and compared to baseline values. The results revealed significant differences between the coronal consonants and the control (/h/) in up to 4 preceding syllables. Formant differences in non-adjacent syllables were limited to F3 (lower for /r/ than the other consonants) and were attenuated in stressed syllables. Non-adjacent LTAS differences were overall more robust, but primarily differentiated between coronals and the non-coronal /h/. Overall, /r/ showed the greatest effect on non-adjacent preceding vowels, followed by /l/, and then by /d/. The formant and LTAS methods appear to capture somewhat different, yet equally important aspects of local and long-range coarticulation. The LTAS findings suggest that higher-frequency information, while generally disregarded for speech, may contain significant coarticulatory information.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Coarticulation is pervasive in speech production, as articulatory gestures of individual phonemes overlap in time and space, producing complex patterns of contextual variation. This process commonly involves immediately adjacent phonemes, as, for example, fronting of a velar stop closure before a front vowel (key vs. coo), or backing of the alveolar constriction of /u/ after a retroflex approximant (burn vs. bun). Less pervasive, but nonetheless well-attested is coarticulation involving non-adjacent phonemes, especially vowels coarticulating in backness, height, and rounding across an intervening consonant (Öhman, 1966; Recasens, 1989; Magen, 1997; Farnetani, 1990). Many models of speech production (e.g. Keating, 1990; Browman & Goldstein, 1990; Perkell & Matthies, 1992) were designed to incorporate both local and non-local VCV coarticulatory effects. For example, with respect to the latter, Articulatory Phonology (Browman & Goldstein, 1989, 1990) assumes that consonant gestures (such as the tongue body movement to make a closure with the velum for /k/) are super-imposed on the inherently longer vowel gestures. Thus, despite being non-adjacent segmentally, the vowels in VCV immediately adjacent at the corresponding articulatory tiers; as they partly overlap with each other (and with the intervening the consonant), their constriction degree and location values ‘blend’ with each other, producing expected patterns of VCV coarticulation.

Less clear is how the models of speech production can capture coarticulation of vowels and consonants to non-adjacent consonants. Such cases of coarticulation appear to be less common and involve certain classes of sounds. Bladon & Nolan (1977), who collected video-fluoroscopy data from 8 speakers of British English (RP) found that the normally apical alveolar stops /t/ and /d/ were produced as laminal when followed or preceded by the laminal fricatives /s/ and /z/ across a vowel (e.g. in deserve, sedan, does, and sat). More temporally extensive coarticulatory effects have been observed with English liquids, the ‘dark’ rhotic (retroflex) approximant /l/ and the ‘clear’ lateral approximant /l/. Kelly & Local (1986) observed that these consonants spread their dark/clear resonances to non-adjacent syllables within a word (e.g. Henry vs. Henley). Tunley (1999) examined the acoustic effect /l/ on adjacent and non-adjacent vowels in CVCV and JCVC sequences produced by three speakers of British English. She found that unstressed non-adjacent vowels (as well as stressed or unstressed adjacent vowels) were clearly influenced by pairs with word-in itial /g/ and /l/ (electropalatography (EPG), and acoustic analysis. Data from three speakers of British English producing minimal pairs with word-initial /l/ and /l/ (reap vs. leap, lobe vs. robe, etc.) in a carrier phrase Have you uttered a ___ at home? revealed that significant differences between the target liquids were detected up to two syllables before and two syllables after (e.g. in [...twood zip shut...]). Specifically, consonants and vowels in these syllables had a higher and more back tongue body, greater lip rounding, and lower F3 when the utterance included /l/.

West (2000) investigated the extent of /l/ and /l/ coarticulation using a combination of articulography (EMA), electropalatography (EPG), and acoustic analysis. Data from three speakers of British English producing minimal pairs with word-initial /l/ and /l/ (reap vs. leap, lobe vs. robe, etc.) in a carrier phrase Have you uttered a ___ at home? revealed that significant differences between the target liquids were detected up to two syllables before and two syllables after (e.g. in [...twood zip shut...]). Specifically, consonants and vowels in these syllables had a higher and more back tongue body, greater lip rounding, and lower F3 when the utterance included /l/.

Heid & Hawkins (2000) conducted an acoustic investigation of the extent and various prosodic and segmental factors on the liquid resonance spreading. One of their findings was that the effect of the target consonant (/l/ vs. /l/ vs. /h/) was detectable even 4 syllables away, as for example, in the word heard [haddr] in the sentence We heard it might be (a) ram/lamb/ham. Interestingly, the closer to the target consonant word might did not show an effect when it was stressed, yet apparently allowing the [i]/[I]-coloring to pass through to the preceding unstressed syllables.

Overall, the findings of small, yet persistent, long-range coarticulatory effects involving coronal consonants are interesting and theoretically important, as they suggest that our current understanding of speech production mechanisms is incomplete, and models of speech production require some revisions. If coarticulation is exclusively a by-product of gestural overlap, and consonant gestures have a specified (and relatively short) extent in time, how do we explain articulatory and acoustic differences manifested several syllables away from the target gesture?

It should be noted that the research on long-range coarticulation is still at its earliest stage, and most, if not all, work done so far is on British English. This is in contrast with a range of studies of vowel-to-vowel (VCV) coarticulation in a variety of dialects and languages. While substantial amount of articulatory and acoustic work has been done on North American /l/ and /l/ (e.g. MRI imaging and modeling: Alwan, Narayanan, & Haker, 1997; Narayanan, Alwan, & Haker, 1997; Espy-Wilson, Narayanan, Alwan, & Boyce, 1997), very few have dealt explicitly with the temporal extent and coarticulatory influences involving these consonants (e.g. Boyce & Espy-Wilson, 1997). Further, it is not clear from the studies reviewed above whether the so-called ‘liquid resonance’ differences are equally due to both /l/ and /l/, or primarily due to the latter. As these studies did not seem to examine other coronal consonants, it is not clear to what extent the effects are specific to the liquid approximants and are not shared by other tongue tip/blade articulations, such as /t, d, n/.
The current study investigates the extent /l/ and /g/ coarticulation in Canadian English, with the goal to replicate some of the results previously reported in the literature on British English. Specifically, we are interested in how far (how many syllables away) acoustic differences between the target consonants are detectible in a sentence. To determine differences in the extent and amount of coarticulation between liquid and non-liquid coronals, and between coronals and non-coronals, we also include alveolar stop /d/ and ‘placeless’ /h/ as target consonants. To quantify coarticulation we use both the traditional formant (F1, F2, and F3 frequencies) method and a novel method calculating long-term average spectra (LTAS) of vowels. The latter method is used because coarticulatory effects may affect the entire spectrum of a vowel, and some differences could be overlooked by the formant-only analysis.

METHODS

Participants

Fourteen young adults from the University of Toronto community were recruited for this experiment (5 male, 10 female). Participants had no self-reported speech, language, hearing or neurological disorders, and spoke Canadian English as a first language. Participants were paid $10/hour for their participation. This experiment was approved by the University of Toronto Research Ethics Board.

Materials and Procedure

There were four target sentences: We thought it might be a ram/ham/lamb/dam. These sentences were chosen as the closest approximation to those in Heid & Hawkins (2000), who found a 4-syllable long anticipatory coarticulatory effects (see above).

Target sentences were interspersed with distractor sentences and presented on a computer screen. Each sentence remained in display for fifteen seconds. Subjects were instructed to repeat the sentence at a normal to fast speaking rate for the duration of its display. Sentences were presented in a fixed random order. The entire set of sentences was repeated three times. Subjects typically produced between 10 and 15 repetitions of each sentence per trial, yielding approximately 30-45 repetitions of each sentence per subject, depending on their overall speaking rate.

Audio was recorded in a sound-attenuating booth with a solid-state Sound Devices 722 recorder and a DPA 4011 cardioid microphone at 44.1 kHz in the Department of Linguistics Phonetics Lab.

Analysis

Preprocessing

Vowel nuclei were manually annotated with Praat (Boersma & Weenink, 2012). Since the liquids /l/ and /g/ do not usually have abrupt boundaries in the spectrogram, as /d/ does, the boundaries of the liquids were taken to be the approximations of F2 and F3 with F1. Since there is typically no obvious break between the vowel nuclei in the sequence ‘...be a...’ they were treated as one vowel, and annotated as such.

Quantifying coarticulation: Formant Method

Two different methods for measuring long-range coarticulation were employed. For the first method, the median formant value for F1-F3 was extracted from each vowel nucleus. The median was chosen over the mean to avoid distortions arising from formant tracking errors. Values for each formant were normalized within subject.

Quantifying coarticulation: LTAS Method

This method draws on the intuition that long-range coarticulation may have an impact on the global shape of the spectrum that may be missed from simply examining the local maxima of the spectrum (i.e. formants). In order to quantify the effects of long-range coarticulation, a measure of spectral distance from baseline was calculated as follows.

For all vowel nuclei, a 100 Hz bandwidth long-term average spectrum (LTAS) was extracted in Praat. Each LTAS was normalized so that its total sum was 1. This was done so that each vowel has the same amount of total energy, just different distributions (i.e. any differences observed will not be the results of different speaking
intensity). Within each subject, a principal components analysis (PCA) was computed over all the vowel nuclei to reduce the number of dimensions of the LTAS. Since many frequency bins strongly co-vary with one another, this allowed a large number of redundant components to be discarded. For each subject, the first $n$ components which accounted for 95% of the variance were retained, and the remainder discarded. The average number of components retained was 103, less than half of the original size of the LTAS which had 221 bins. The range was 80-130 components. These reduced LTAS (rLTAS) spectral representations served as the basis of the acoustical analysis.

Next, baseline-centroids were computed. In the baseline condition (where the target word was *ham*, i.e. the target consonant had no oral place of articulation), within each subject, an average rLTAS was calculated for every vowel nuclei by computing the average of each of the rLTAS components over all trials within subject. This creates an array of 44 $n$-dimensional centroids: ([14 subjects] x [6 syllables]).

Finally, spectral distance, $D$, from baseline was calculated as follows. For some given trial, the appropriate centroid for each syllable was found in the array, and the Euclidean distance between the rLTAS of that vowel and the centroid was computed, giving a single measure of acoustical distance for each syllable which could be submitted to a Repeated Measures ANOVA.

It should here be clear why a distinct set of centroids is calculated for all subjects. If rLTAS centroids are calculated across subjects, then, necessarily, some subjects' production will, on average, be closer to the overall average, and it will incorrectly appear that they produce less long-range coarticulation compared to other subjects, since what we would be measuring is the subject's deviation from the overall average, and not the long-term spectral effects of /a/, /u/ or /d/.

### RESULTS

#### Formant Method

For all three formants, some groups failed the Kolmogorov-Smirnov test for normality. Therefore, the non-parametric, multivariate Scheirer-Ray-Hare test was used. The factors were Syllable (6 levels) and Target Consonant (4 levels). Significant effects were found for Syllable on F1 [F(5, 0.006), $H=0.0073$, $p=0$], F2 [F(5, 0.0042), $H=0.006$, $p=0$] and F3 [F(5, 752), $H=0.002$, $p=0$]. Significant effects were found for Target Consonant on F1 [F(3, 57.09), $H=42.37$, $p<0.5*10^{-35}$] and F3 [F(3, 371), $H=695$, $p=0$]. Significant interactions was found for F1 [F(15, 13.91), $H=51.63$, $p<0.5*10^{-35}$] and F3 [F(15, 1292), $H=715$, $p=0$].

Figure 1 shows the changes in formant space over the course of the sentence by target consonant. Each node in the graph shows the mean F2 and F3 value for that vowel and target consonant. It can be seen that, while there are early divergences in the vowels' locations in formant space as early as the second and third words of the sentence, it is relatively small compared to the dramatic divergence which occurs around the target consonant.

Post-hoc tests showed no significant differences in F1 or F2 for target consonant except for locally, and so these formants are ignored in further discussion. In the third formant, however, post-hoc tests found significant differences for non-local vowel formants. Figure 2 shows the mean values of normalized F3 for each vowel, in each target consonant condition. Error bars indicate standard error. Asterisks indicate that a significant difference was found between the two groups' means using a post-hoc Tukey test where alpha=0.05.

It can be seen that F3 of [æ] immediately following the target consonant (L) shows a 4-way significant difference depending on the nature of that consonant. The magnitude of the difference is the greatest between /a/ (low F3) and the three non-rhotic consonants (high F3). The latter difference is also very robust in the immediately preceding syllable (*be a, vowels [i]/*), while the differences within the non-rhotics are non-significant. No significant differences were found in the word *might* ([æ][i]), although F3 in the /a/ condition was on average lower than for the other consonant conditions. In the next two words, *it* ([i]) and *thought* ([æ][i]), F3 differences are significant for the pair /a/ vs. /h/ (lower for the former), and non-significant for the other pairs. Finally, the most distant word from the target consonant, *we* ([i]), shows no significant differences in F3. Notably, F3 is still on average slightly lower for the /a/ condition compared to the other consonant conditions.
FIGURE 1. Mean normalized F2 and F3 values for each vowel and target consonant condition.

FIGURE 2. Mean normalized F3 values for each vowel (in *we, thought, it, might, be a,* and the target word (L)) and each target consonant condition (/h/, /d/, /l/, /r/).
FIGURE 3. The negative logarithm of acoustical distance (D) for each vowel (in we, thought, it, might, be a, and the target word (L)) and each target consonant condition (/h/, /d/, /l/, /g/).

FIGURE 4. The root mean squared distance between the LTAS centroids and individual trials (Delta) calculated for each vowel (in we, thought, it, might, be a, and the target word (L)) and each target consonant condition (/h/, /d/, /l/, /g/).

LTAS Method

Since the measure of acoustical distance, D, can only have positive values, it was log-transformed to avoid possible floor effects. Kolmogorov-Smirnov tests failed to reject the null hypothesis that the data followed a normal distribution for all groups with alpha=0.001, and so here a conventional 2-way ANOVA was used. Significant differences were found for Syllable [F(5, 218.4), p=0] and Target Consonant [F(3, 235), p=0]; a significant interaction was also found [F(15, 34.1), p=0]. Figure 3 shows the mean values of the negative logarithm of D for each vowel in each consonant condition. Error bars indicate standard error. Asterisks indicate a significant difference was found between group means using post-hoc Tukey test with alpha=0.05.
As seen the figure, the vowel of target word (L) shows a 4-way significant differentiation of target consonant conditions. The immediately preceding vowel (in be a) distinguishes among all the consonant conditions, except /i/ vs. /l/. The difference between the coronals and non-coronal /h/ is particularly robust. These general coronal/non-coronal differences persist in the preceding two syllables (thought and it), while becoming somewhat more limited (/l/ vs. /h/ in it and /i/, /l/ vs. /h/ in thought). No significant differences were found in the first word of the sentence, we; however, average values here are still higher for /h/ than for the coronal consonants.

To give some support to the notion that the LTAS method provides more results because it examines deviations in the spectrum as a whole, rather than at single points, the root mean squared distance between the LTAS centroids and individual trials was calculated over each LTAS bin. This measure is called Delta. The results are shown in figure 4. For example, in the word thought, the majority of the difference between the baseline centroid and individual trials occurs at relatively low frequencies – around 2 kHz. However, in the word it, much of the difference between the baseline centroids and individual trials occurs in the 10 kHz range.

**DISCUSSION AND CONCLUSION**

The goal of this study was to replicate earlier findings of long-range coarticulatory effects involving liquids /l/ and /s/. As all previous work on the subject was done on British English, to our knowledge, this is the first systematic investigation of the effect in North American English. Consistently with Heid & Hawkins (2000), whose study was based on a single RP speaker, our data from 14 Canadian English speakers revealed significant acoustics differences up to four (or technically 5) syllables away from the target consonant. Recall that similar, but temporally more limited (to 2 syllables) effects were reported in other studies on British English liquids (Kelly & Local, 1986; Tunley, 1999; West, 2000). We also replicated Heid & Hawkins’ (2000) finding that stressed syllables are resistant to /s/ resonance spreading, yet do not ‘block’ the spread of /s/-quality to preceding syllables. This was manifested in non-significant F3 differences in the closer to the target word might, but significant differences in the preceding words thought and it.

The acoustic analysis in this study supplemented the traditional formant method (used in the previous works reviewed above) with the long-term average spectrum (LTAS) method. The formant method, as mentioned earlier, did not produce statistically significant results for F1 and F2 formants in non-adjacent syllables. This is different from the previous studies, where F2 values were found to differentiate /l/ from /s/ and /h/ (Tunley, 1999; West, 2000). This could be attributed to the somewhat different resonances of /l/ in Canadian English (darker, and thus more similar to /i/). On the other hand, the formant analysis yielded temporally extensive differences in F3 (4 preceding syllables). Notably, F3 was on average lower in the /s/ condition in all six analyzed syllables of the utterance. Not all these differences, however, reached the significance level, being limited to the pair /s/ vs. /h/ in non-adjacent syllables.

Compared to the formant method, the LTAS method has yielded overall more robust and consistent significant differences in 5 out of 6 analyzed syllables. The difference between /h/ and other consonant conditions steadily increased and gradually involved more contrasts as the vowels were closer to the target consonant (see Figure 3; cf. Figure 1). This is probably the result of looking at the entire spectrum, as evidenced by the Delta measure. Additionally, significant differences were found for the word might which appeared to be ‘transparent’ to long-range coarticulation in the formant analysis, a result which was also found by Heid & Hawkins’ (2000) analysis. The patterning of coronal consonants as a class in the LTAS results perhaps suggests that the method is sensitive to spectral properties associated with more general articulator movements such as the tongue front raising and fronting, rather than the more phoneme-specific characteristics such as the creation of sublingual cavity for /s/ reflected in lower F3. Overall, the LTAS results demonstrate that higher-frequency information, while generally disregarded for speech, may actually contain significant non-local coarticulatory information. The two methods, thus, appear to capture somewhat different, yet equally important, aspects of local and long-range coarticulation.

All three coronal consonants examined in the study showed coarticulatory effects on preceding adjacent and non-adjacent vowels; the consonants, however, differed in the relative extent of their influences, roughly in the order /s/ > /l/ > /d/ (based on both formant and LTAS results). While this suggests that coronal consonants as a class are capable of exerting long-range coarticulatory effects (cf. Bladon & Nolan, 1977; Gafos, 1999), it also shows that liquids, and particularly /s/ are stronger ‘triggers’ of coarticulation, consistently with the theory of liquid ‘resonance spreading’ (Kelly & Local, 1986).

To conclude, the results of the current study provide new evidence for long-range coarticulation involving English consonants. Together with previous work on the phenomenon, these findings raise questions about the mechanisms of speech production and planning, having implications for the development on new phonetic models.
ACKNOWLEDGMENTS

This work was funded by a Standard Grant #416-2010-0959 from Social Sciences and Humanities Council of Canada to Alexei Kochetov.

REFERENCES


