

# An acoustic investigation of the [ATR] feature effect on vowel-to-vowel coarticulation

Christina Orphanidou, Greg Kochanski and John Coleman

Phonetics Laboratory, University of Oxford, UK

christina.orphanidou@phon.ox.ac.uk, greg.kochanski@phon.ox.ac.uk, john.coleman@phon.ox.ac.uk

## Abstract

We report quantitative measurements of coarticulation of the Advanced Tongue Root (ATR) feature in neighbouring and non-neighbouring vowels. Native speakers of Southern British English produced utterances; we matched pairs that were identical except for a contrast of [+ATR] vs. [-ATR] on one “transmitter” vowel. From recordings, we computed an acoustic representation at the centres of the vowels, measuring the difference in pronunciation associated with the contrast in [ATR] value. We observed a consistent coarticulatory effect two syllables after the transmitter when it was a low vowel (i.e. the / $\Lambda$ / vs. /a:/ and / $\partial$ / vs. / $\sigma$ :/ pairs), but little or no effect when the transmitter was a high vowel (i.e. the /i:/ vs. /i/ and / $\upsilon$ / vs. /u:/ pairs). Interestingly, we observed more dramatic coarticulation on the next-nearest neighbour than on the vowel following the transmitter.

**Index Terms:** coarticulation, ATR, articulatory target.

## 1. Introduction

Phonological theory holds that words are constructed out of distinctive features and presumably [1] that these features have observable roles in speech production and perception. We test this connection in the context of speech production by searching for consistent articulatory and coarticulatory effects of those features. This study reports on the Advanced Tongue Root (ATR) feature. The [ATR] feature contrasts tense (+) and lax (–) vowels; it has been proposed as a phonological feature by [6] but its existence in English has been challenged by [7]. If [ATR] is a valid feature, it should also have a straightforward association with certain acoustic changes.

Coarticulation can provide a test of feature-hood. In the standard “feature spreading” view, common in the phonological literature, features are either specified or not (c.f. [2]) and the effects of specified features spread out across unspecified regions [26]. However, feature spreading terminates where a feature is specified, so if [ATR] is a feature, there should be no phonological effects on the far side of a vowel that specifies [ATR]. As a result, phonological coarticulation of [ATR] should be blocked by any vowel that specifies the contrary value of [ATR]. We work within three-syllable regions of a three- or four-syllable utterance, looking at forward-propagating “carry-over” coarticulation.

Phonetic coarticulation models typically describe speech in terms of articulatory gestures or targets. Coarticulation is then described in terms of overlap of two gestures [25], as a result of inertial or mechanical limitations of the articulators [16], or a planning process to reach phonologically specified soft targets [15, 17].

In this study we investigated the local and coarticulatory

acoustic correlates of the [ATR] feature. We conducted a systematic survey of the strength of [ATR]-driven vowel-to-vowel coarticulation for both adjacent and across-vowel cases using speech samples that differ only in the [ATR] feature. By keeping all other phonemes the same, any effect we observe on subsequent vowels can be attributed to the [ATR] difference. Furthermore, if there is an articulatory target related to the [ATR] feature, the corresponding acoustic properties should display relatively little variability across different environments and the coarticulatory effects should be fairly consistent.

While much past research on vowel-to-vowel coarticulation focused on nearest-neighbour effects [15, 4, 18, 19], non-nearest-neighbour coarticulation has also received a lot of attention [10, 11, 3, 14, 9, 5, 8], with most studies concluding that there is a non-nearest-neighbour vowel-to-vowel coarticulatory effect which is attenuated by intervening phonemes that strongly specify a target. Our study aims to investigate the existence of coarticulatory effects over a variety of intervening phones and using a large number of subjects in order to reduce the effect of inter-subject variability. Our study differs from most coarticulation experiments in that it searches for the coarticulatory effect of a minimal contrast between vowels that differ only in one feature.

## 2. Experimental Methods

Twenty-seven speakers (fifteen male and twelve female) took part in the experiment. All subjects were native speakers of Southern British English, aged 19–34, and all were students or members of staff of Oxford University. They were seated in a noise-insulated room and were asked to read out phrases which appeared on a computer screen. The signal was recorded with a noise-cancelling microphone and was digitized at 16 bit precision with a 44.1 kHz sampling rate. In an initial task, subjects read out fifty long, phonetically rich sentences which were used for another experiment. For the present experiment, each subject read out an average of 456 sentences, randomly taken from a pool of 408 sentences and presented in random order. (Four percent of the sentences were randomly chosen to be read four times.) A total of 12760 sentences were recorded but this paper analyses only the replicated ones.

### 2.1. Speech Material

The text consisted of  $(CV)CV^tCV^rCV^dC(VC)$  tri- and tetra-syllabic utterances. All the utterances were English phrases, mostly noun phrases, chosen to be the tail of a plausible English sentence. In the utterances,  $V^r$  (the “resistor” vowel) was chosen from the set of /i:/, /i/, / $\Lambda$ /, /a:/, / $\partial$ /, / $\sigma$ :/, / $\upsilon$ /, /u:/, /e/, /a/ and / $\partial$ / and  $V^d$  (the “detector” vowel) was always a / $\partial$ /.  $V^t$  is

the “transmitter” vowel; each sentence was paired with another sentence which was identical except for the  $[\pm\text{ATR}]$  feature of  $V^t$ . The vowel pairs tested were /i:/ vs. /ɪ/ (e.g. “beach hunter” vs. “bitch hunter”), /ʌ/ vs. /a:/ (“it hums operas” vs. “it harms operas”), /ɒ/ vs. /ɔ:/ (e.g. “they stock lemurs” vs. “they stalk lemurs”) and /ʊ/ vs. /u:/ (e.g. “pull to the thing” vs. “pool to the thing”). The combinations of phonemes before and after  $V^t$  define the different “contexts” within which the phoneme under investigation is studied. The set of contexts was chosen to cover a wide variety of features. A total of 53 contexts were used for the /ʌ/ vs. /a:/ pair, 33 for the /ʊ/ vs. /u:/ pair, 48 for the /ɒ/ vs. /ɔ:/ pair and 224 for the /i:/ vs. /ɪ/ pair. Speech samples were segmented using the HTK Speech Recognition toolkit [22] based on a transcription derived from UNISYN<sup>1</sup>[12, 24].

### 3. Analysis

#### 3.1. Acoustic Description Vector

The computation of the acoustic description vectors follows [20] closely; it represents the sounds within 45 ms of the vowel’s midpoint. Most of the vector is computed from a “perceptual spectrum” [21], which is a power spectrum, collected in 0.7 erb-wide bins, then raised to the 1/3 power to approximate the perceived specific loudness of a sound. The vector contains these specific loudnesses averaged over a 60 ms window, edge detectors that respond to changes in the spectral power on a 45 ms time scale, and a spectral entropy measure. It also contains two components not derived from the specific loudnesses: a voicing estimator and a measure of dissonance. The dissonance measure follows [13] and corresponds to the summed modulation of the cochlear vibration from 7 Hz up to the effective bandwidth of the cochlea at each point.

We used these acoustic description vectors to train a classifier to distinguish between sounds that are phonologically the same vs. different. Specifically, it was given differences between pairs of acoustic description vectors and trained to distinguish between two classes: those pairs that were obtained from equivalent points in phrases spoken from the same text, and those pairs that were not from equivalent points.

Then, we took a step beyond [20]. We converted the classifier that was developed there into an approximate, acoustically-based measurement of phonological distance. This was done by recognising that its Equation 2 is equivalent to a simple, unweighted Euclidean distance in a different coordinate system, one defined by

$$\vec{\alpha} \cdot \vec{\alpha}^T = \vec{a}M\vec{a}^T, \quad (1)$$

where  $M$  is defined in [20] and describes the classifier,  $\vec{a}$  is an acoustic description vector, and  $\vec{\alpha}$  is the corresponding vector in the new coordinate system. Explicitly,  $\vec{\alpha} = \vec{a}PD^{1/2}$ , where  $PDP^T = M$  is the eigenvalue decomposition of  $M$ . We can safely drop small eigenvalues to obtain a new 16-dimensional coordinate system which has the useful property that Euclidean distances in it are a good approximation to phonological distance. This means that all components are equally important and correlations have been removed. Using these vectors, the classifier can correctly distinguish pairs of phonologically identical sounds from pairs chosen from nonequivalent positions

82% of the time<sup>2</sup>. Here, we used differences between these transformed acoustic description vectors.

#### 3.2. Measuring the effect of $[\pm\text{ATR}]$

In order to investigate the effect of the  $[\pm\text{ATR}]$  feature we calculated the *difference vectors* between the *average* acoustic description vectors for the transmitter ( $V^t$ ), resistor ( $V^r$ ) and detector ( $V^d$ ) vowels, subtracting vectors from identical contexts and varying  $V^t$  by  $[\pm\text{ATR}]$ . If an articulatory target exists for the  $[\text{ATR}]$  feature then its coarticulatory effect should be consistent, i.e. the difference vectors for the resistor and detector vowels should display relatively little variability. In other words, the difference vectors should be pointing in approximately the same direction and should be of similar length. Figure 1 helps visualise this. In it, we project the average acoustic description

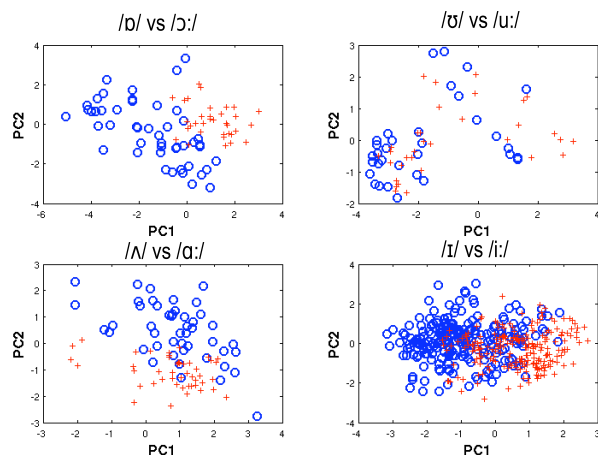


Figure 1: *Acoustic description vectors of the transmitter vowels in four  $[\pm\text{ATR}]$  pairs. The figure shows plots of the first two principal components of the acoustic description vectors averaged over each context. There are 224 contexts for the /i:/ vs. /ɪ/ pair, 48 for the /ɒ/ vs. /ɔ:/ pair, 33 for the /ʊ/ vs. /u:/ pair and 53 for the /ʌ/ vs. /a:/ pair. Circles denote the  $[-\text{ATR}]$  average vectors and crosses the  $[\text{ATR}]$  average vectors. We can observe that there is a higher degree of overlap for the /i:/ vs. /ɪ/ and /ʊ/ vs. /u:/ than for the other two, indicating that the two vowels are spectrally more similar compared to the /ɒ/ vs. /ɔ:/ and /ʌ/ vs. /a:/ which are spectrally more distinct.*

vector for every context onto a plane using Principal Component Analysis [23] by taking the two largest principal components. We calculated this for the 4 pairs of  $[\text{ATR}]$  and  $[-\text{ATR}]$  vowels. The extent of overlap of the two clouds of data points reflects the degree of spectral similarity of the two vowels i.e. the less overlap there is, the more likely they are to have clearly defined articulatory targets. The number of points in each one of the four plots equals the number of contexts included in the experiment for each pair.

We also measured the distance between  $[\text{ATR}]$  and  $[-\text{ATR}]$  contexts for the transmitter, detector and resistor vow-

<sup>1</sup>Unisyn is a software package for predicting pronunciation in a variety of English dialects. It is based on a lexicon and includes transformation rules.

<sup>2</sup>The phoneme classification accuracy is several percent higher because “inequivalent” positions will sometimes randomly sample the same phoneme, if the phoneme occurs more than once in the phrase. Such accidental phone matches occur about 5% of the time.

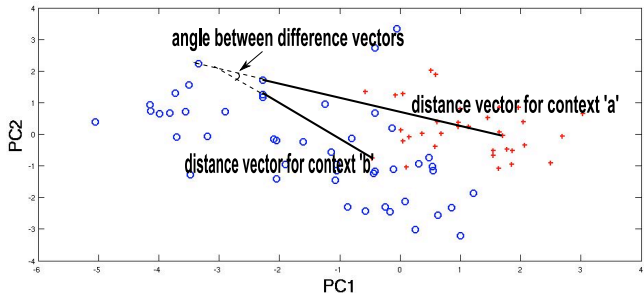


Figure 2: Acoustic description vectors for the transmitter vowels of the /v/ vs. /ɔ:/ pair. The principal components of the acoustic description vectors averaged over the context are plotted against each other for the two vowels. The solid lines indicate the difference vectors between the averages of the /v/ and /ɔ:/ vowels for two specific contexts. The angle between the two vectors indicates the strength of the (co-)articulatory effect relative to the normal utterance-to-utterance variation. If the angle is near  $\pi/2$  rad, the (co-)articulatory effect caused by [ $\pm$ ATR] is much smaller than variation; if the angle is small, the (co-)articulatory effect is consistent and much larger than the variation.

els, by computing the length (norm) of the difference vectors. This distance allows us to measure of the size of coarticulation relative to variation or the direct articulatory effect of changing [ATR]. Measurements for the transmitter vowel provide a baseline against which to interpret the angle and distance measurements for the other two.

Figure 2 illustrates how the angles between the difference vectors are calculated, showing data from the /v/ vs. /ɔ:/ pair. For every context the difference vectors are calculated between the transmitter, detector and resistor vowels (where the transmitter is the [ $\pm$ ATR] pair). After all the three difference vectors are calculated for every different context the angles between them (for different contexts) are computed in order to see if they have a consistent direction. If the angles between the vectors are small, the direction is consistent; it corresponds to the [+ATR] contexts forming a cluster that is well separated from the [-ATR] contexts.

#### 4. Results

Figure 3 shows the means of the angles between the difference vectors for the transmitter, resistor and detector vowels.

If the angle is near  $\pi/2$  rad, the (co-)articulatory effect caused by [ $\pm$ ATR] is much smaller than variation; if the angle is near zero, the (co-)articulatory effect is consistent and much larger than the variation. We can see that the low vowel pairs /v/ vs. /ɔ:/ and /ʌ/ vs. /ɑ:/ show a clear coarticulatory effect whereas the high vowel pairs /v/ vs. /u:/ and /i:/ vs. /ɪ/, show a smaller effect. Table 1 shows the distance values of the different vectors. As expected, the distances between the transmitter vowels are much greater than the distances between the detector and resistor vowels (which are the same vowels). Furthermore, there is some consistency in the size of the coarticulatory effect between the different pairs.

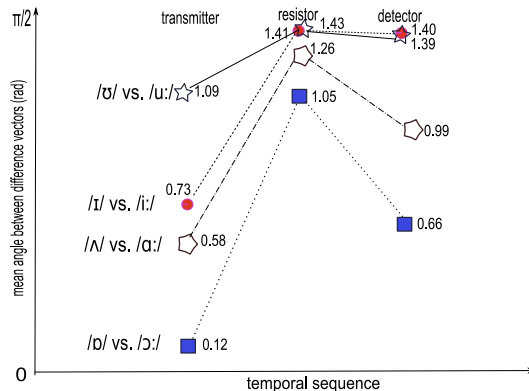


Figure 3: Mean angles between difference vectors for the 4 [ $\pm$ ATR] pairs, in rad. We can see that the low vowels /v/ vs. /ɔ:/ and /ʌ/ vs. /ɑ:/ pairs show a large coarticulatory effect (with mean angles smaller than  $\pi/2$ ) whereas the high vowels /v/ vs. /u:/ and /i:/ vs. /ɪ/ don't. Somewhat surprisingly there is a more consistent coarticulatory effect on the detector rather than the resistor vowels.

Table 1: Table of distances between average acoustic description vectors for different contexts.

transmitter pair		
/v/ vs. /ɔ:/	detector distance	0.5
	transmitter distance	2.8
	resistor distance	0.4
/u:/ and /v/	detector distance	1.1
	transmitter distance	2.7
	resistor distance	0.9
/i:/ vs. /ɪ/	detector distance	0.6
	transmitter distance	2.1
	resistor distance	0.6
/ɑ:/ and /ʌ/	detector distance	0.6
	transmitter distance	2.5
	resistor distance	0.7

#### 5. Conclusion

We conducted an experiment in order to investigate the strength of the [ATR] feature on vowel-to-vowel forward coarticulation using 27 speakers of Southern British English. We found, first, that our Southern British English speakers make strong [ATR] distinctions on the low vowels. The distinctions were weaker and less reliable on high vowels. Similarly, across many different contexts, [ATR] will coarticulate across a vowel and modify a following schwa. This is a strong effect when the [ $\pm$ ATR] difference causing the coarticulation is on a low vowel, but it causes little or no coarticulation when a high vowel.<sup>3</sup> In addition, we have shown an across-vowel coarticulatory effect of a minimal contrast by comparing vowels differing only by one

<sup>3</sup>We note that the difference in coarticulation cannot easily be accounted for by invoking a nonlinear articulatory to acoustic mapping, because the high/low difference is on the vowel causing the coarticulation, not on the schwa where we measure it. The schwa's acoustic properties (and thus presumably its articulatory position) are fairly stable and not strongly affected by whether the transmitter vowel is high or low. Thus, the schwa's articulatory to acoustic mapping should always be fairly stable and the magnitude of the acoustic effect should directly relate to the articulatory effect.

feature. Finally, unexpectedly, we observed stronger coarticulation across a “resistor” vowel onto a schwa than we observed on the resistor vowel itself. We are not aware of any theoretical models that provide a natural explanation for this effect. Mechanical, dynamical systems models generally predict coarticulatory effects that decrease with distance, and phonological feature spreading would presumably be blocked by the resistor vowel.

## 6. Acknowledgements

This research was conducted under grant RES-000-23-1094 from the Economics and Social Research Council; the authors also thank Linacre College for kindly providing accommodation for the conduct of the experiments.

## 7. References

- [1] Ohala, J. J., “Experimental Phonology”, in Goldsmith, J. A. (ed.) *The Handbook of Phonological Theory*, Blackwell, Oxford, 1995, pp. 713–722, ISBN 0-631-18062-1.
- [2] Steriade, D., “Underspecification and Markedness”, *ibid.* pp. 114–174.
- [3] Benguerel A.-P. and Cowan H.A., “Coarticulation of upper lip protrusion in French”, *Phonetica*, 30:41–55, 1974.
- [4] Cohn A. C., “Nasalisation in English”, *Phonology*, 10:43–81, 1993.
- [5] Coleman J., “Discovering the acoustic correlates of phonological contrasts”, *J. Phonetics*, 31:351–372, 2003.
- [6] Halle M. and Stevens K., “On the feature Advanced Tongue Root”. *Quarterly Progress Reports*, MIT Research Lab for Electronics, 94:209–215, 1969. Republished in *From Memory to Speech and Back*, M. Halle, Mouton de Gruyter, 2003.
- [7] Harshman P. and Goldstein L., “Factor analysis of tongue shapes”, *J. Acoustical Soc. America*, 62:693–707, 1977.
- [8] Hawkins S. and Nguyen N., “Influence of syllable-coda voicing on the acoustic properties of syllable-onset // in English”, *J. Phonetics*, 32:199–231, 2004.
- [9] Magen H., “The extent of vowel-to-vowel coarticulation in English”, *J. Phonetics*, 25:187–205, 1997.
- [10] Öhman S. E. G., “Coarticulation in VCV utterances: Spectrographic measures”, *J. Acoustical Society of America*, 39:151–168, 1966.
- [11] Öhman S. E. G., “Numerical models of co-articulation”, *J. Acoustical Society of America*, 41:310–320, 1967.
- [12] Fitt, S. and Isard, S. “Synthesis of regional English using a keyword lexicon,” in *Proceedings: Eurospeech 99* (v. 2), The International Speech Communication Association, (1999) pp. 823–836, [http://homepages.inf.ed.ac.uk/sue/Publications/Fitt\\_1999\\_a.ps](http://homepages.inf.ed.ac.uk/sue/Publications/Fitt_1999_a.ps), (Viewed 11/2006).
- [13] Pierce, J. R., *The Science of Musical Sound*, Scientific American Books, New York (1983), pp. 74–81.
- [14] Slater A. and Coleman J., “Non-segmental analysis and synthesis based on a speech database”, In *Proceedings of the 5th Speech Production Seminar*, pp. 105–108, Seon, Germany, 2000.
- [15] Whalen D. H., “Coarticulation is largely planned”, *J. Phonetics*, 18:3–35, 1990.
- [16] Modarresi G., Sussman H., Lindblom B. and Burlingame E., “An acoustic analysis of the bidirectionality of coarticulation in VCV utterances”, *J. Phonetics*, 32:291–312, 2004.
- [17] Kochanski, G., Shih, C., and Jing, H., “Quantitative measurement of prosodic strength in Mandarin, *Speech Communication* 41(4), 2003, pp. 625–645. doi:10.1016/S0167-6393(03)00100-6
- [18] Cho T., “Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English”, *J. Phonetics*, 32:141–176, 2004.
- [19] Beddor P. S., Harnsberger J. D. and Lindemann S., “Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates”, *J. Phonetics*, 30:591–627, 2002.
- [20] Kochanski, G. and Orphanidou, C., “Testing the Ecological Validity of Repetitive Speech”, In *Proceedings of the XVIth International Congress on Phonetic Sciences*, Saarbrücken, Germany, 6–10 August, 2007. Paper 1632, <http://www.icphs2007.de/conference/Papers/1632/1632.pdf>, viewed 2/2008.
- [21] Bu L. and Chiueh T.-D., “Perceptual speech processing and phonetic feature mapping for robust vowel recognition”, *IEEE Transactions on Speech and Audio Processing*, 8(2):105–114, 2000.
- [22] Cambridge University Engineering Department. *HTK Speech Recognition Toolkit*, January 2007.
- [23] Pearson, K., “On Lines and Planes of Closest Fit to Systems of Points in Space”, *Philosophical Magazine* 2(6):559–572, 1901.
- [24] Fitt, S., “Unisyn Lexicon Release”, <http://www.cstr.ed.ac.uk/projects/unisyn/> (UNISYN.1.1, viewed 4/2006), The Centre for Speech Technology Research, The University of Edinburgh, 2002.
- [25] Fowler C. A. and Saltzman E., “Coordination and Coarticulation in Speech Production”, *Language and Speech*, 36(2,3):171–195, 1993.
- [26] Henke, W. Dynamic articulatory model of speech production using computer simulation. MIT PhD dissertation, 1966.

This paper was presented at the ISCA Tutorial and Research Workshop on Speech Analysis and Processing for Knowledge Discovery, June 4-6, 2008, Aalborg University, Denmark. Dr. Christina Orphanidou presented. It is available at [http://kochanski.org/gpk/papers/2008/isca\\_coart.pdf](http://kochanski.org/gpk/papers/2008/isca_coart.pdf) .

The book of abstracts may be available at [http://kom.aau.dk/project/ITRW/resources/proceedings/isca\\_itrw\\_aalborg\\_2008\\_book\\_of\\_abstracts.pdf](http://kom.aau.dk/project/ITRW/resources/proceedings/isca_itrw_aalborg_2008_book_of_abstracts.pdf) .

## Scientific Topics

- Parametric and nonparametric models
- New all-pole and pole-zero spectral modelling
- Temporal modelling
- Non-spectral processing (group delay etc.)
- Integration of spectral and temporal processing
- Biologically-inspired speech analysis and processing
- Interactions between excitation and vocal tract systems
- Characterization and representation of acoustic phonetic attributes
- Attributed based speaker and spoken language characterization
- Speech analysis and processing for detecting acoustic phonetic attributes
- Language independent aspects of acoustic phonetic attributes detection
- Detection of language-specific acoustic phonetic attributes
- Acoustic to linguistic and acoustic phonetic mapping
- Mapping from acoustic signal to articulator configurations
- Merging of synchronous and asynchronous information
- Other related topics