AN ARTICULATORY MODELING OF ROMANIAN DIPHTHONG ALTERNATIONS

Stefania Marin

Yale University
stefania.marin@yale.edu

ABSTRACT

This paper presents an articulatory modeling of the alternation between Romanian diphthong *ea* and unstressed vowel *e,* starting from the hypothesis that the representation of Romanian diphthongs is that of two vowels synchronously coordinated. Stimuli are created to examine the effect of this synchronous coordination in the absence of stress, and two perceptual experiments show that synchronously coordinated vowels [e] and [a] result in the percept of an [e]-like blended vowel – the same outcome as reported in Romanian phonological alternations.

Keywords: Romanian diphthongs, task dynamics application, Articulatory Phonology

1. INTRODUCTION

Romanian diphthongs *ea* and *oa* need to be phonologically specified since they contrast with single vowels *e* and *o* (1a), with glide-vowel sequences *ja* and *wa* (1b) and with hiatus sequences *e.a* and *o.a* (1c). Furthermore, these diphthongs, always stressed, alternate with unstressed *e/o* (2a).

(1) a. '**sea**.ra EVENING-DEF '**se**.ra GREENHOUSE-DEF
b. '**bea**.ta DRUNK-F-SG '**bja**.ta POOR-F-SG
c. **re**.'**al** REAL '**deal** HILL

(2) a. Alternating roots:
'**sea**.ra EVENING-DEF **se**.'ra.ta EVENING SHOW -DEF
b. Non-alternating roots:
'**se**.ra G.HOUSE-DEF **se**.ri.'ci.ca G.HOUSE (DIM)

Previous experimental evidence ([5]) has shown that unstressed [e] in alternating roots (2a) is significantly different from [e] in non-alternating roots (2b), a difference that could not be explained as a stress effect. This kind of evidence prompted the proposal, framed within an Articulatory Phonology theoretical approach ([1], [2]), that Romanian diphthong alternations are the result of a specific articulation coordination pattern, i.e. diphthongs [ea]/[oa] are two vowels synchronously coordinated. When stress affects the two vowels differently, the result is a stressed diphthong (Figure 1b); when there is no stress effect, the result is a blended vowel [e]/[o] (Figure1a).

The present paper examines the validity of this proposal by attempting an articulatory modeling of Romanian diphthong alternations. To this end, artificial stimuli were created by using an articulatory synthesis model, described in Section2, and by manipulating the model parameters in line with the proposal made here, as detailed in Section 3. Then, the stimuli thus created were used in two perceptual experiments, presented in Section 4.

Figure 1a: Synchronous coordination of vowel gestures

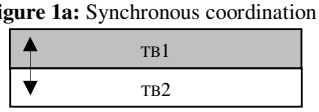
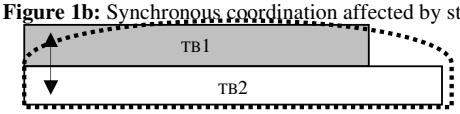


Figure 1b: Synchronous coordination affected by stress



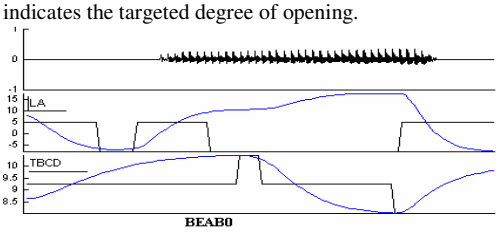
2. DESCRIPTION OF THE MODEL

TADA (Task Dynamics Application) is a computational system developed over several decades at Haskins Laboratories, Inc. to test the hypotheses put forward by dynamic speech production models such as Articulatory Phonology ([1], [2]). The model consists of a Linguistic Gestural Model ([1]) that generates gestural scores to be computed by a Task Dynamic model of inter-articulator coordination ([9]). Then, the Task Dynamic model generates articulator trajectories that are used by a vocal tract articulatory synthesizer ([8]) to compute sound. (cf. Figure 2 for modeled made-up [be.ab]).

This version of the model implements inter-gestural coordination specifications, such as synchronous or sequential coordination, by overlapping gestures, i.e. by specifying relative inter-gestural timing. This specification is what can be controlled in this model to test the vowel coordination hypothesis proposed for analyzing

Romanian diphthongs and the observed phonological alternations and contrasts.

Figure 2: Gestural representation for made-up [be.ab] produced by TADA. The input is specified in ARPABET (BEH)(AAB), with brackets representing syllabification. The boxes indicate gestural activation and the curves the generated tract variable movement. Height of the boxes indicates the targeted degree of opening.

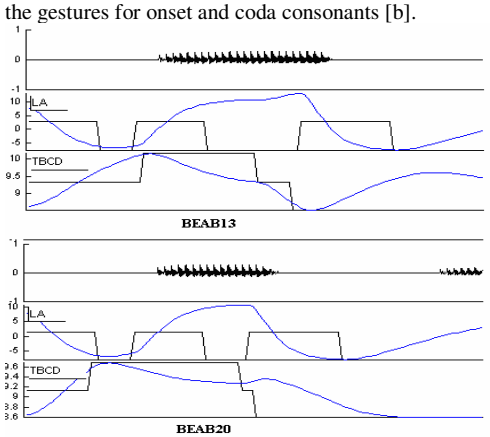


3. STIMULI

3.1. Stimuli design

Using TADA, the stimuli were constructed starting from the gestural specification for two vowels syllabified in hiatus (the default of the model). Subsequent stimuli were then manipulated by stepwise increasing the overlap between the two vowels, with the underlying assumption that greater overlap is the result of synchronous coordination between articulatory gestures, while less overlap/more sequentiality of gestures is the result of sequential or other types of coordination.

Figure 3: Stimuli gestural scores and tract variables. The relevant vowel tier is TBCD (Tongue Body Constriction Degree), while LA (Lip Aperture) captures the gestures for onset and coda consonants [b].



The base stimulus, BEAB0 [be.ab] (Figure 2), was created with the articulatory gestures for [e] and [a] sequential in time. For the other stimuli, gestural activation for vowel [a] was shifted a specific number of temporal frames earlier relative to the activation interval of [a] in BEAB0, while gestural activation for [e] was maintained constant. Thus, for the next stimulus created, BEAB5, the gesture for [a] started and ended 5 frames earlier relative to the starting and ending point for [a] in BEAB0, and so on for all 13 stimuli created: BEAB0, BEAB5, BEAB10-20, with the identifying number representing the number of frames that [a] was shifted relative to activation of [a] in BEAB0. For stimulus BEAB20, the gesture for vowel [a] starts 9 frames after the gesture for vowel [e] starts, but it ends 2 frames before gesture [e] ends, and for this reason, this stimulus was considered as the end of the series, i.e. this represented full overlap between the two gestures. Figure 3 gives the gestural scores and tract variables for stimuli BEAB13 and BEAB20. While manipulating overlap, the duration of the stimuli was also affected, resulting in *EA* duration range from 381ms (BEAB0) to 173ms (BEAB20).

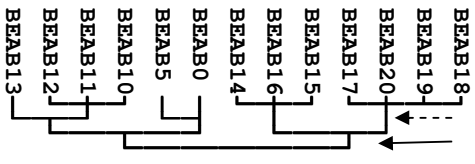
3.2. Stimuli acoustic analysis

A Hierarchical Cluster Analysis (SPSS 12.00) was used to explore possible grouping of the synthesized stimuli based on their acoustic properties: duration of vowel portion and formant frequency. This procedure is an exploratory method that attempts to identify groups of cases based on selected characteristics (variables) and it uses an algorithm that starts with each case in a separate cluster and combines clusters until only one is left.

Stimulus duration and formant frequency were measured using Praat 4.1.18 speech analysis software. For formant measuring, the vowel portion was manually isolated and then F1 and F2 at 25%, 50% and 75% landmarks in the interval were automatically detected.

The results of Hierarchical Cluster Analysis are presented in Figure 4 (the same configuration is obtained if duration of vowel portion is also included in the analysis). At the two-group level (solid arrow), stimuli BEAB14-20 and BEAB0-13 end up grouped together, and at the second level of analysis (dotted arrow), stimuli BEAB17-20, BEAB14-16, BEAB10-13 and BEAB0-5 end up grouped together.

Figure 4: Hierarchical Cluster Analysis. *Variables:* F1 and F2 at 25%, 50% and 75% landmark.



Quantity differences were an expected result of overlap. However, hierarchical cluster analysis suggests that the stimuli are also qualitatively different from each other in a meaningful way since use of formant frequency variables leads to classifying these stimuli in several groups.

4. PERCEPTION EXPERIMENTS

4.1. Subjects and procedure

Ten subjects (1 female, 9 male), whose native language was Romanian participated in both perceptual experiments (Experiments 1 and 2). DMDX 3.1.1.3 software was used to present the auditory stimuli over a set of Sennheiser PX100 headphones and the answers were recorded using different keys on the computer's keyboard (no reaction time was measured).

4.2. Experiment 1

4.2.1. Experimental design

This experiment utilized a forced-choice identification procedure, in which the 10 listeners heard the experimental item and had to decide whether the stimulus heard was a) part of two words, b) a diphthong, or c) a single vowel. The stimuli were presented in random order, and each stimulus was presented five times during the experiment. All the stimuli created were used.

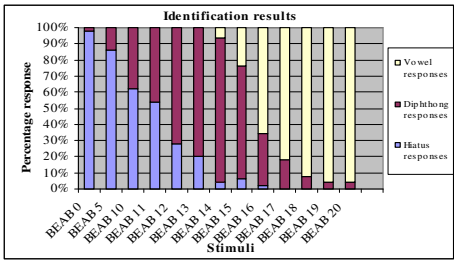
4.2. 2. Results

Individual responses were averaged for each subject across the five responses for a given stimulus, and then these individual percentage responses were further averaged across speakers. These results are plotted in Figure 5.

The listeners progressed from identifying the base-line stimulus BEAB0 as a hiatus word to identifying the more overlapped stimuli, starting with BEAB12, as diphthong-words and finally to identifying the most overlapped stimuli, starting with BEAB16, as *e*-words, supporting thus the "blending" hypothesis, that truly simultaneous *e* and *a* would result into what is perceived by native speakers as an *e*.

In term of inter-subject response variability, the most extreme stimuli (the baseline and the most overlapped) were ascribed to the same category (hiatus or vowel respectively) in a consistent manner, while for the middle stimuli there was slightly more variability between subjects' responses. Stimulus BEAB10 was identified as a hiatus by a majority of speakers, with quite a lot of diphthong responses as well, but crucially with no single vowel responses. At the other end, the most overlapped stimulus BEAB20 was consistently identified as vowel [e]. Hence these two stimuli were used as extremes for a forced-categorization of the intermediate items in a discrimination experiment, presented in the following section.

Figure 5: Results of identification task.



4.3. Experiment 2

4.3.1. Experimental design

The same subjects and general procedure as in Experiment 1 were used in this discrimination AXB-type experiment. Each stimulus in this experiment consisted of one of the stimuli BEAB11 to BEAB19 presented flanked by BEAB10 (identified mostly as a hiatus stimulus [e.a] in the previous experiment, and never identified as vowel [e]) and BEAB20 (identified as vowel [e] in the previous experiment). The listeners had to decide whether the middle sound was more like the sound preceding or following it. Each stimulus was presented 6 times, in random order, either preceded or followed by the [e]-like stimulus.

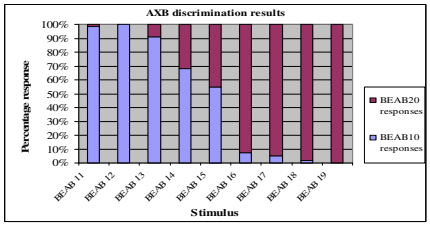
4.3.2. Results and discussion

As in the previous experiment, individual responses were averaged for each subject across the six responses for a given stimulus, and then these individual percentage responses were further averaged across speakers. The results are plotted in Figure 6.

The results of this experiment are consistent with the results in the identification task. Starting with stimulus BEAB16, listeners consistently identify the items as more [e]-like (BEAB20) than hiatus-like (BEAB10), which is the same point at which in the previous experiment listeners started to identify the

stimulus as a vowel. However, while in this discrimination task, listeners classify stimulus BEAB16 as a vowel at a 93% rate, they only classified it as a vowel 66% of the time in the identification task. In spite of this difference, which was not unexpected given the particularities of the two tasks, it is important to observe that both tasks have the same cut-off point at which overlap results into a vowel-like precept for a majority of the subjects. Thus, the discrimination experiment comes to confirm the blending hypothesis, that the more overlapped items (i.e. those with synchronous coordination between vowels [e] and [a]) result into an [e]-like precept.

Figure 6: Results of discrimination task.



It must be noted that there was no stress manipulation in these stimuli, and only overlap of the two vowels was varied. A better articulatory model that would also manipulate stress is therefore desirable to fully model the alternation conditions observed in Romanian phonology, and this is the object of a future study. However, even with this limitation, what the stimuli used in these two experiments showed was that two synchronously coordinated vowels [e] and [a] result in the precept of an [e]-like vowel in the absence of any other external factors such as stress.

One question to address is whether the observed vowel percept was indeed due to a qualitative difference and not just an effect of the different duration of the stimuli. It might be that listeners categorized stimuli starting with BEAB16 as a vowel because they were shorter than a typical Romanian diphthong, rather than because they sounded as an [e]-like vowel.

Duration of Romanian diphthong *ea* is reported with a range between 115ms - 230ms, averaging 120 ms ([4]) to 187ms ([7]), and that of *e* with a range between 70ms-112ms, averaging 92ms ([3]). While, as expected, there is variation, it must be noted that naturally produced diphthongs are shorter than the vowel portion durations of most of the artificial stimuli reported here. Even the shortest stimulus BEAB20, with vowel interval duration of 173ms, is well within the range reported for naturally produced diphthong [ea].

This suggests that duration alone of these stimuli could not have induced a single vowel percept just because these stimuli were too short to qualify as diphthongs. It can be concluded therefore that the perceptual results were triggered by a qualitative difference, corroborated with inherent shortening resulting from more overlap.

5. CONCLUSION

The articulatory modeling presented in this paper has demonstrated that two synchronously (i.e. fully overlapped) coordinated vowels [e] and [a] result in the precept of an [e]-like vowel in the absence of any other external factors such as stress. This supports the proposal that complex nucleus effects, illustrated in (1) and (2), can be explained by a specific pattern of gestural organization – synchronous coupling, and by lawful consequences predicted by such coupling (Figure 1a-b).

On a larger scale, this proposal addresses the notion of syllables and what syllables are, and adding to work done on complex onset and coda effects ([1], [2], [6]), it comes to complete, from the complex nucleus effects perspective, the hypothesis that what defines syllables is not an arbitrary hierarchy grouping segments, but rather a specific mode of coordination between gestures.

REFERENCES

[1] Browman, C. P, Goldstein, L. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299-320.

[2] Browman, C. P, Goldstein, L. 2000. Competing constraints on inter-gestural coordination and self-organization of phonological structures. *Bulletin de la Communication. Parlée* 5, 25-34.

[3] Burileanu, D. 2002. Basic research and implementation decisions for a text-to-speech synthesis system in Romanian. *Int. J. of Speech Technology* 5, 211-225.

[4] Chitoran, I. 2002. A perception-production study of Romanian diphthongs and glide-vowel sequences. *J. of the IPA* 32: 2.

[5] Marin, S. 2005. Complex Nuclei in Articulatory Phonology: The Case of Romanian Diphthongs. In: Gess, R., Rubin, E. (eds), *Selected papers of the 34th LSRL*. Amsterdam, Philadelphia: John Benjamins, 161-177.

[6] Nam, H, Saltzman, E. 2003 A competitive, coupled oscillator model of syllable structure. *Proc. 15th ICPhS* Barcelona, 2253–2256.

[7] Rosetti,A, Avram,A, Cocian,C, Ghitu,G, Suteu, V, Zamfirescu,I, Cocenai,S. 1955. Experimental studies on Romanian diphthongs. *Studii şi Cercetări Lingvistice* 6, 7-27.

[8] Rubin, P., Baer, T., Mermelstein, P. 1981. An articulatory synthesizer for perceptual research. *JASA*. 70, 321-328.

[9] Saltzman, E. L., Munhall, K.G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382.