Romanian ‘blended’ vowels:
A production model of incomplete neutralization

STEFANIA MARIN

Abstract

This study proposes a production model for the incomplete acoustic neutralization between underived and derived /e/ in Romanian. Using the articulatory-based synthesizer TADA, underived /e/ was modeled with a single articulatory gesture, while derived /e/ was modeled as a ‘blending’ between two vocalic gestures timed synchronously (similar to the diphthong /ea/ with which it alternates). A comparison of the acoustic properties of modeled and naturally produced stimuli showed that underived /e/ tokens were acoustically similar to modeled underived /e/ and that naturally produced derived /e/ tokens were similar to modeled ‘blended’ /e/. This result supports the hypothesis that derived /e/ is the result of a blending between two vowel gestures, and that the observed incomplete acoustic neutralization between underived and derived /e/ in Romanian is the result of different articulatory mechanisms.

Introduction

Past research has shown that certain derived consonants and vowels are phonetically different from their underived equivalents, with which they would be expected to be homophonous based on transcriptions. For example, final devoiced stops in German, Dutch, Polish, or Catalan have been shown to be phonetically different from underlying voiceless stops, although both are customarily transcribed as voiceless (cf. Warner, Jongman, Sereno, & Kemps, 2004, for a review). Likewise, although coda liquids in Puerto Rican Spanish have been described as merging to a common category, recent research has shown that they exhibit subtle spectral differences which betray the underlying category to which they correspond (Simonet, Rohena-

ISSN 1645-4537

-Madrazo, & Paz, 2008). A further case is that of coda obstruents in Eastern Andalusian Spanish: these obstruents are all produced as aspiration, but their durational properties are different depending on whether they correspond to an underlying /s/, /p/, or /k/ (Bishop, 2007). Such fine phonetic differences are not limited to consonants: Hungarian ‘transparent’ vowels have been shown to differ in their articulation, betraying their origin as either back or front vowels (Benus & Gafos, 2007). In Romanian, vowel /e/ alternating with diphthong /ea/ has been reported to be produced slightly, but significantly, more centralized compared to underived /e/ (Marin, 2005). These examples share the property that apparently merged/neutralized categories maintain subtle phonetic (articulatory and/or acoustic) differences, beyond expected contextual or sociolinguistic variation.

Such incomplete neutralization phenomena have proven theoretically challenging for traditional segmental phonological approaches which assume that no information of the phonological processes or paradigmatic relations should be available at the production level, and that therefore segments represented by the same abstract symbolic category should be realized the same, regardless of whether they are derived or not (see Port, 1996 for a discussion). These cases emphasize the need for a linguistic model that would allow for differences at the representational/phonological level to be reflected at the implementation/production level (cf. Warner et al., 2004). In this context, the goal of the current paper is to examine one such case of incomplete neutralization, namely that observed in Romanian between certain derived and non-derived /e/ vowels, and to propose a production model accounting for this phenomenon. In the specific model adopted, differences at the planning (phonological) level are reflected directly at the execution level by assuming an identity between the units of representation and those of production (cf. Browman & Goldstein, 1989, 1992). Under this approach, if the units of planning are the same as the units of production, then a derived form may be produced differently as a direct result of its different representation at the planning level.

Romanian phonology distinguishes between derived /e/, alternating with diphthong /ea/ (1a), and underived /e/, in non-alternating roots (1b). Previous preliminary research has shown that derived /e/ is realized acoustically as more centralized compared to underived /e/ (Marin, 2005). This significant difference between underived and derived /e/ was consistent across different lexical items tested, and was not explainable as a stress effect: underived stressed /e/ in Romanian was not spectrally different from underived unstressed /e/. Interestingly, derived /e/ was not acoustically different from the /e/ portion of the diphthong it alternates with, suggesting that its properties were similar to those of an /e/ vowel co-produced with vowel /a/. These observations suggested that derived /e/’s acoustic properties could be the result of a simultaneous, ‘blended’ co-production of both vowels /e/ and /a/, analogous to the diphthong with which it alternates (for a discussion of the difference between derived /e/ and diphthong /ea/ in production, see Marin, 2005). Under this hypothesis, underived /e/ is assumed to be produced as a canonical vowel /e/ with a single gesture (in the sense of Articulatory Phonology, cf. Browman & Goldstein, 1989, 1992), while derived /e/ is the result of a co-production ('blending') between two vowel gestures. As a result, derived /e/'s acoustic difference from underived /e/, and its similarity to /e/ produced in the context of vowel /a/ (in the diphthong) are the result of its bi-gestural representation, directly reflected in production.

(1) a. Alternating roots:
['te̲a̲.mə] 'fear' [te̲.mə.'tor] 'fearful'
['dʒe̲a̲m] 'window' [dʒe̲.mu.'lets] 'window (Diminutive)'

b. Non-alternating roots:
['te̲.mə] 'homework' [te̲.mi.'tʃi.kə] 'homework-
-Diminutive'
['dʒe̲m] 'jam' [dʒe̲.mi.'ʃor] 'jam (Diminutive)'

In a perceptual experiment (Marin, 2007), a stimulus modeled articulatorily with the gestures for vowels /e/ and /a/ fully overlapped was identified by 10 listeners as vowel /e/, providing supporting evidence for the hypothesis that derived /e/ may be the result of synchronously articulated vowels /e/ and /a/, while stimuli where vowels /e/ and /a/ overlapped 90% or less were identified as diphthong /ea/. In the current paper, the hypothesis that derived /e/ is different from underived /e/ as a result of its bi-gestural representation is further tested by comparing the acoustic outputs of tokens produced by native speakers with stimuli modeled articulatorily. Specifically, under the proposed hypothesis, naturally produced underived /e/ tokens should be acoustically similar to stimuli modeled with only one articulatory target (that for vowel /e/), while naturally produced derived /e/ tokens should be acoustically similar to stimuli modeled with a blending of two articulatory targets. If the acoustic properties of modeled and naturally produced stimuli turn out to be similar, it may be inferred that their articulatory properties are also similar, and thus the articulatory configurations of natural vowels could be inferred from the known articulatory configurations of the modeled ones.

Method

The model

The computational model used in the current study is the Task-Dynamic Application (TADA), an articulator-based system developed at Haskins Laboratories to test hypotheses formulated within dynamical speech production models such as Articulatory Phonology (Browman & Goldstein, 1990; Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984; Goldstein, Byrd, & Saltzman, 2006; Nam, Goldstein, & Proctor, n.d.; Saltzman & Munhall, 1989). TADA generates speech outputs on the basis of dynamical specifications of gestures (modeled as critically damped oscillators) and the coupling relations between them (as schematized in Figure 1 for the words of interest for this paper). Specific coupling relations between gestures at the planning level are assumed to correspond to specific timing patterns at the production level: in-phase coupling results in synchronous articulatory timing, while anti-phase coupling results in sequential timing.

Modeled articulatory trajectories (exemplified in Figure 2) are computed to satisfy these specifications, i.e. the gestures' targets and their timing. On the basis of these trajectories, vocal tract shapes and area functions are determined, which serve as input for the generation of sound using the pseudo-articulatory synthesizer HLSyn (Hanson & Stevens, 2002). The acoustic output thus generated on the basis of known articulatory configurations can then be compared to speaker-produced acoustic outputs. Rather than assuming symbolic units at the planning/phonological level, and specific articulatory targets at the execution level, the model uses the same units throughout, thus ensuring transparency between the levels.

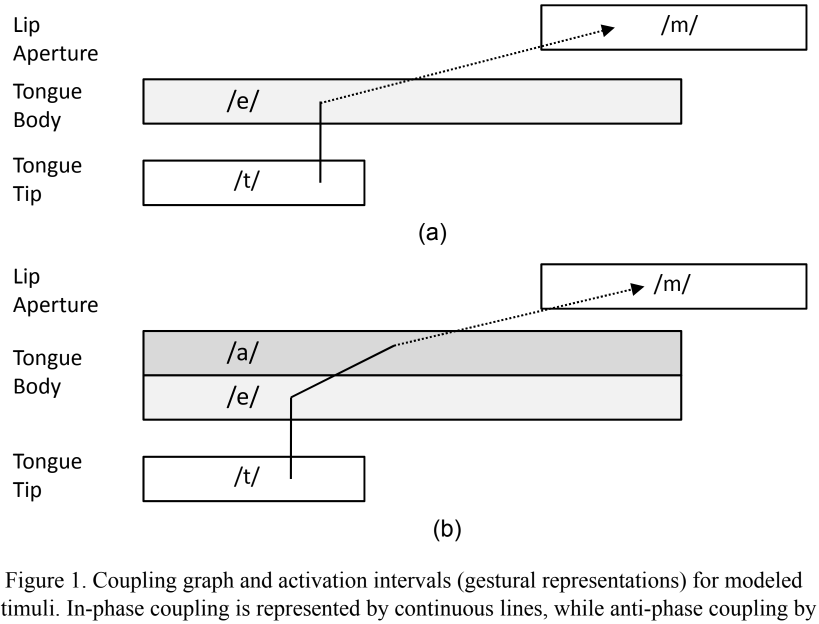

Figure 1. Coupling graph and activation intervals (gestural representations) for modeled stimuli. In-phase coupling is represented by continuous lines, while anti-phase coupling by dashed arrows. The gestures in curly brackets are not shown. (a) Modeled underived /e/ in ['te̲.m{ə}]'homework: /t/ and /e/ are coupled in-phase, /e/ and /m/ are coupled anti-phase; (b) Modeled blended /e/ in [te̲.m{ə.'tor}]'the evening show': /e/ and /a/ are coupled in-phase to each other and respectively in-phase to /t/ and anti-phase to /m/.

Stimuli

For the current experiment, underived and derived /e/ stimuli (['te̲.mə] *'homework'* – [te̲.mə.'tor] *'fearful'*) were modeled starting from the gestural representations illustrated in Figure 1. Thus, underived /e/ in ['te̲.mə] was modeled with a single vowel gesture (for /e/), while derived /e/ in [te̲.mə.'tor] was modeled with two vowel gestures (for /e/ and for /a/) coupled in-phase. The latter modeled stimulus is referred to as modeled 'blended' /e/. The default articulatory parameters of TADA for the specific vowels were used. Thus, underived /e/ was defined by a palatal tongue body constriction (95 degrees on an imaginary arc along the palate going from 0 degrees at the teeth to 180 degrees at the pharyngeal region), and a wide constriction degree (tongue body 10.5 mm away from the palate). 'Blended' /e/ was defined as the simultaneous (in-phase) production of vowel /e/ (with the same specifications as those used for underived /e/) and of vowel /a/ (defined with a tongue body pharyngeal constriction at 180 degrees and 11mm away from the palate). This specification would result in a vocal tract with a tongue body constriction and degree representing the blending (averaging) between /e/ and /a/. The resulting articulatory trajectories and acoustic signals are shown in Figure 2.

The same stimuli were also produced by nine native speakers of Romanian (six female). In addition, the diphthong word ['te̲a̲.mə] *'fear'* was recorded to serve as reference for derived /e/, with which it alternates. Two control-/e/ pairs were also recorded: a stress-control pair (['be̲.ri.le] *'the beers'* – [be̲.'ri.kə] *'beer-Diminutive'*), and a word-length-control pair (['te̲.me] *'homeworks'* – ['te̲.me.le] *'the homeworks'*). These control pairs were not recorded for one of the male speakers. In the stress-control pair, non-alternating stressed /e/ was compared to non-alternating unstressed /e/; in the word-length-control pair, non-alternating /e/ in a two-syllable word was compared to /e/ in a three-syllable word. The control pairs were selected so that within each pair the consonantal context of the target vowel as well as the vowel in the following syllable would be the same in the two conditions. Given lexical limitations, no exact vowel or consonantal contexts could be used across pairs. Each stimulus was embedded in a constant carrier phrase, and it was read ten times in random order, in blocks combined with additional words used for other experiments. The stimuli were presented on a computer screen one at a time, and speakers were instructed to read them at a self-selected casual speaking rate. The recordings were made in Romania, in a quiet room, using a digital recorder and a Behringer Ultravoice XM8500 microphone. All the recordings were sampled at 22.05 kHz.

Acoustic analysis

The acoustic signal of both natural productions and modeled stimuli were analyzed using Praat speech analysis software (Boersma & Weenink, n.d.).

The vocalic interval of interest was manually labeled from the onset to the offset of the vowel-specific formant contours, and formant frequencies for five formants were automatically calculated using Praat's short-term spectral analysis function, with the following parameters: Burg-algorithm-computed LPC coefficients, a 25ms Gaussian window, with a frame shift of 5 ms, and a pre-emphasis of 50 Hz. The formants were calculated within a 5000 Hz range for the male speakers and for the modeled stimuli, and within a 5500 Hz for the female speakers. Formant trajectories calculated using these parameters were visually inspected for accuracy. The formant frequency values were converted within Praat to the auditory Bark scale using the formula in Schroeder, Atal, & Hall (1979), as an intrinsic vowel normalization method (Harrington & Cassidy, 1999).

The first two formant frequency values at the mid-point of the measured interval were extracted for analysis. The mid-point was chosen under the assumption that this is usually the time point around which achievement of target for the vowels occurs, and it is reasonably the time point least influenced by context (Harrington & Cassidy, 1999; Van Son & Pols, 1990). It is therefore likely the best temporal landmark for comparing the acoustic properties of underived and derived /e/. The Bark distance between the two formants (F2-F1) was used for all statistical analyses as a further method of gender normalization (cf. Syrdal & Gopal, 1986, who suggest that using Bark differences between formants significantly reduces gender variability between speakers, in comparison to Bark or Hertz single formant measurements). Vowel duration was also computed on the basis of the labeled onsets and offsets.

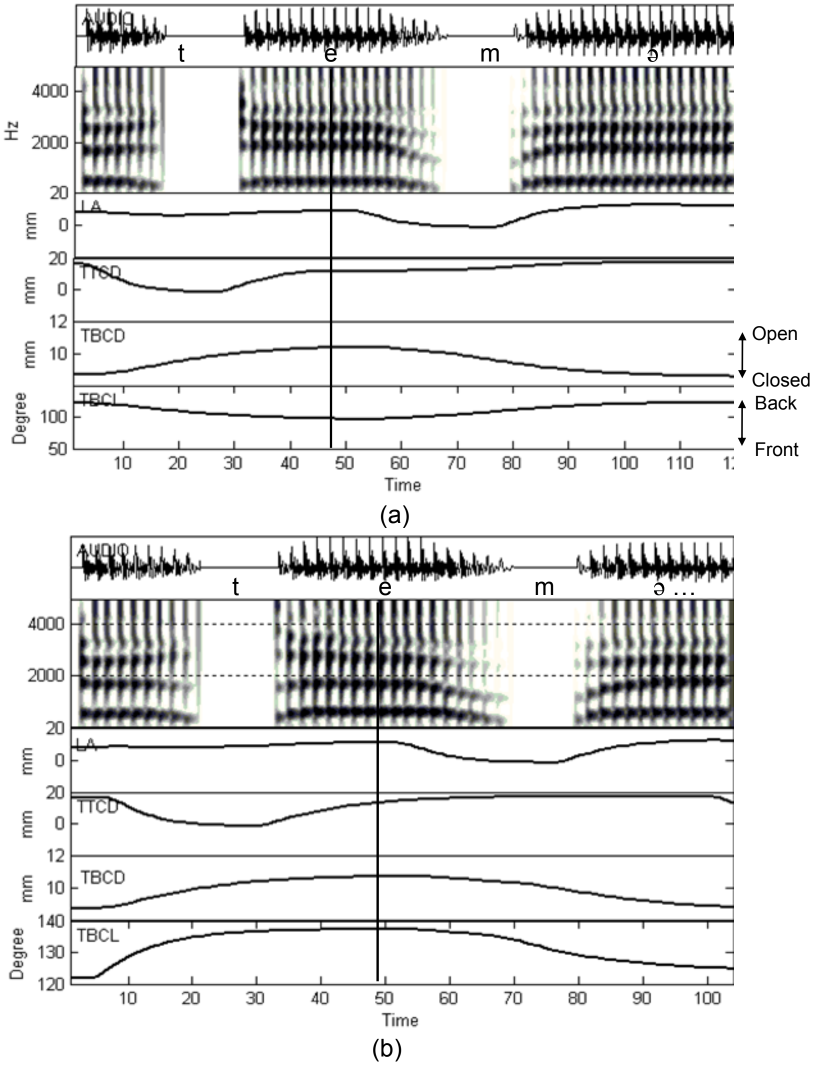

Figure 2. Articulatory trajectories computed to satisfy the representations shown in Figure 1, and corresponding acoustic outputs: Lip Aperture (LA) for /m/, Tongue Tip Constriction Degree (TTCD) for /t/, Tongue Body Constriction Degree (TBCD) and Tongue Body Constriction Location (TBCL) for the vowels of interest. (a) Modeled underived /e/ in ['te̲.mə]’homework; (b) Modeled blended /e/ in [te̲.mə.{'tor}]’the evening show’. The gestures in curly brackets are not shown (the gestures and acoustic signal for [ə] are only partially shown). The vertical continuous lines show the approximate location used for the acoustic measure.

Acoustic similarity

To quantify the degree of acoustic similarity between natural and modeled stimuli, an adaptation of the procedure employed by Harrington (2006) was used for determining relative proximity in the acoustic domain between different vowels. Specifically, the Euclidean distance (E) from each natural stimulus to every modeled stimulus was calculated on the basis of the F2-F1 values of each token at the acoustic mid-point. Thus, for each naturally-produced stimulus, be it ['te.mə] or [te.mə.'tor], two Euclidean distance values were calculated: one to modeled ['te.mə] ($E_{[\text{'te.mə}]}$) and one to modeled [te.mə.'tor] ($E_{[\text{te.mə.'tor}]}$). To further determine whether a naturally produced underived or derived /e/ was closer to modeled underived or blended /e/, relative proximity indices (P) were calculated by subtracting the token's Euclidean distance to modeled blended /e/ from the Euclidean distance to modeled underived /e/ ($P_{[\text{'te.mə}]/[\text{te.mə.'tor}]} = E_{[\text{'te.mə}]} - E_{[\text{te.mə.'tor}]}$). By this measure, negative values indicate that a natural token is closer acoustically to modeled underived /e/, and positive values that it is closer to modeled blended /e/. Individual token proximity indices were averaged across tokens of the same word by the same speaker, such that each speaker contributed one averaged proximity index (P_{average}) for the word ['te.mə], and one for the word [te.mə.'tor].

The general prediction using this measure was that if natural and modeled categories were acoustically similar, then naturally produced categories should be closer to the respective modeled categories. Thus, the category closer to modeled underived /e/ should be naturally produced underived /e/, and the category closer to modeled blended /e/ should be naturally produced derived ('blended') /e/.

On the other hand, if natural and modeled tokens of a given category were not acoustically similar, we would expect no consistent pattern of similarity between natural and modeled categories. Under this alternate hypothesis, the natural category closer to a given modeled stimulus should not necessarily match the category of the model (e.g., the category closer to modeled [te.mə.'tor] could be naturally produced underived /e/ rather than naturally produced derived /e/), or two categories could be indistinguishably close to the same modeled stimulus (e.g., natural underived and derived /e/ could be both indistinguishably close to, for example, modeled ['te.mə]).

Results

Acoustic analysis of naturally-produced stimuli

Visual inspection of the stimuli showed that although alternating with a diphthong, derived /e/ was produced as a monophthong, confirming traditional descriptions and orthographic conventions (cf. Figure 3 showing representative examples of ['tea.mə] and [te.mə.'tor]). The acoustic properties of the naturally-produced stimuli are plotted in Figure 4. Derived /e/ is characterized by lower

F2 and slightly higher F1 values than underived /e/, but markedly higher F2 and lower F1 values than the diphthong, whose midpoint showed /a/-typical acoustics (consistent with visual observations of the stimuli).

For the statistical analysis, the F2-F1 values at mid-point were averaged for each experimental item across multiple repetitions by the same speaker. The matched-samples t-tests conducted, summarized in Table 1, showed that derived and underived /e/ differed significantly on this acoustic measure, with a lower F2-F1 Bark value for derived /e/ in [te̠.mə.'tor] (*M* = 6.34, *SD* = 0.72), compared to underived /e/ in ['te̠.mə] (*M* = 7.03, *SD* = 0.56). The direction of the difference suggests that derived /e/ was produced more centralized in comparison to underived /e/. This centralization was realized mainly on the F2 dimension, as salient from Figure 4. Given the obvious difference in the F1xF2 dimension between the diphthong stimulus and all other stimuli, no statistical tests were performed to quantitatively compare the diphthong and the other stimuli.

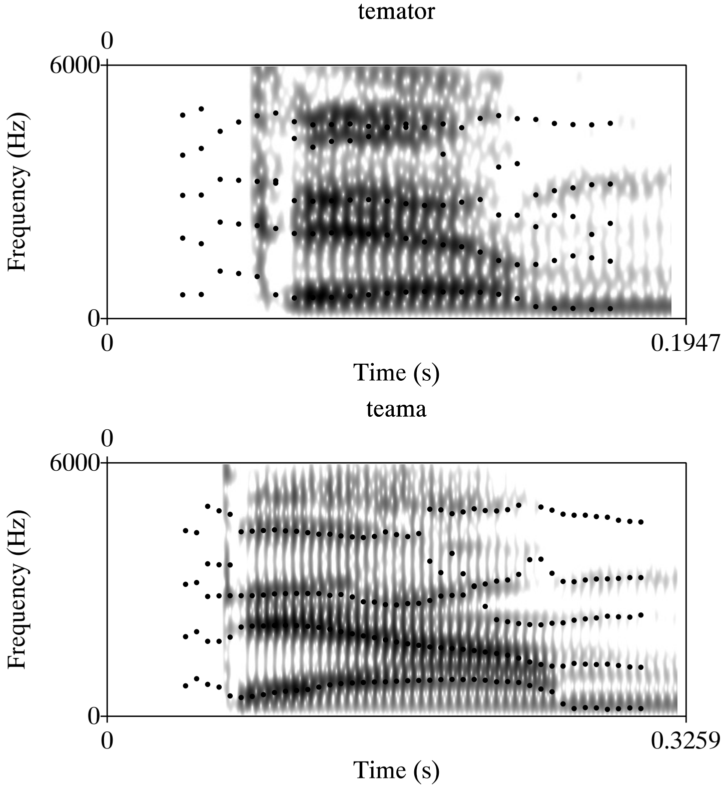

Figure 3. Spectrograms of derived /e/ in [te.m{ə.'tor}] 'fearful' (top) and diphthong /ea/ in ['tea.m{ə}] 'fear' (bottom), as produced by a female speaker. The sounds in curly brackets are not shown.

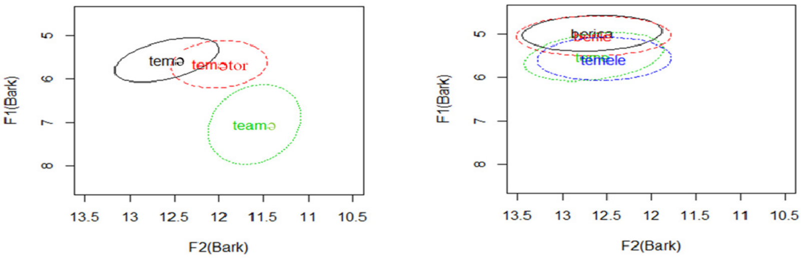

Figure 4. Acoustic characteristics of experimental (left) and control (right) stimuli in the F1xF2 dimension. Stimulus names are positioned to the centroid of each ellipsis. For the stress control and word length control pairs respectively, the two ellipses (and centroids) are on top of each other.

On the other hand, no significant acoustic difference was observed between stressed and unstressed /e/ (['be̲.ri.le]: *M*=7.61, *SD* = 0.97; [be̲.'ri.kə]: *M* = 7.68, *SD* = 0.84), or as a function of word length (['te̲.me]: *M*=7.12, *SD* = 0.81; ['te̲.me.le]: *M*=6.96, *SD* = 0.88). Indeed, their distributions overlap each other (Figure 4). The observed difference between underived and derived /e/ cannot therefore be due to stress and/or word-length effects, since neither stress shift nor an additional syllable affected F2-F1 values significantly in the control pairs. Furthermore, derived and non-derived /e/ are both represented by the same symbol in writing – Roman alphabet letter *'e'* with no diacritics, so the difference between the two /e/-types cannot be attributed to the influence of orthography on production.

Vowel duration was affected by both stress and word length, to the effect that derived /e/ – in unstressed position and part of a three-syllable word, was the shortest; on the other hand, the diphthongs had the greatest duration (cf. Table 2).

Table 1. Statistical results for the matched-samples t-tests conducted on averaged-across-repetitions F2-F1 values at mid-point. The vowel tested is underlined. Effect size (Cohen's *d*) was calculated on the basis of standard deviation values for the groups (Dunlop, Cortina, Vaslow, & Burke, 1996). For the experimental pair, the comparison is between underived /e/ in ['te̲.mə] and derived /e/ in [te̲.mə.'tor].

	Comparison	Two-tailed matched samples t-tests
Stress effect	[be̲.'ri.kə] – ['be̲.ri.le]	*t*(8) = 0.52, *p* = .62, *d* = 0.077
Word length effect	['te̲.me] – ['te̲.me.le]	*t*(8) = 2.17, *p* = .062, *d* = 0.19
Experimental pair	['te̲.mə] – [te̲.mə.'tor]	*t*(8) = 3.50, *p* = .008, *d* = 1.05

Table 2. Means, standard deviations and results for matched-samples t-tests for the vowel duration measure. The vowel tested is underlined. Descriptive statistics are in milliseconds.

	Comparison	Mean (SD)	Matched samples t-tests
Stress effect	[be̲.'ri.kə] ['be̲.ri.le]	105 (21) 120 (33)	*t*(8) = 3.24, *p* = .012
Word length effect	['te̲.me] ['te̲.me.le]	95 (25) 78 (17)	*t*(8) = 5.45, *p* = .001
Experimental pair	['te̲.mə] [te̲.mə.'tor]	93 (23) 59 (14)	*t*(8) = 7.48, *p* < .001
Diphthong	['te̲a̲.mə] ['te̲.mə]	126 (32) 93 (23)	*t*(8) = 8.87, *p* < .001

Comparison between naturally-produced and modeled stimuli

A comparison of the mean formant values of natural tokens with the formant values of modeled stimuli showed that the values for modeled and natural stimuli were quite similar qualitatively, with more centralized formants for derived than for underived /e/ (Table 3).

Table 3. F1 and F2 values measured at vowel acoustic mid-point in Bark for modeled and naturally produced derived and underived /e/ words. Values for modeled items are from a single stimulus; values for the naturally produced items are means across nine subjects.

	Modeled tokens		Natural tokens	
Word	F1	F2	F1	F2
['te̲.mə]	4.87	12.59	5.56	12.59
[te̲.mə.'tor]	5.83	11.73	5.66	11.99

An analysis of the averaged proximity indices ($P_{average}$) quantifying the acoustic similarity between naturally produced and modeled categories indicated that there was an overall match in category between natural and modeled stimuli (Figure 5). Thus, the stimuli acoustically closer to modeled ['te.mə] were naturally produced ['te.mə] words, with a negative mean and median of proximity index $P_{average}$ (*M* = -0.61, *SD* = 0.78); the stimuli acoustically closer to modeled [te.mə.'tor] were naturally produced [te.mə.'tor] words, with a positive mean and median (*M* = 0.48, *SD* = 0.87). A matched-samples t-test carried out on averaged proximity indices confirmed that naturally produced categories differed significantly in terms of their proximity to modeled tokens (*t*(8) = 4.58, *p* = 0.002¸ *d* = 1.32). This analysis showed that naturally produced underived /e/ tokens were significantly closer acoustically to the stimulus modeled with a single gesture for /e/, than to the stimulus modeled as a ‘blending’ between /e/ and /a/. Likewise, derived /e/ productions were more similar acoustically to the bi-gesturally modeled ‘blended’ /e/ than to the mono-gestural model.

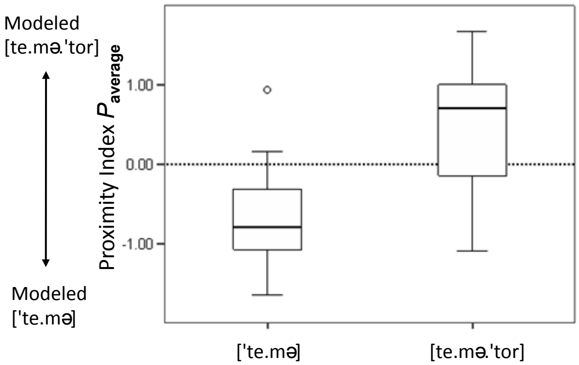

Figure 5. Box plots for the averaged proximity indices $P_{average}$ of naturally produced tokens to modeled tokens.

Discussion

The acoustic analysis of naturally-produced stimuli showed that underived and derived /e/ differed significantly, with derived /e/ being realized more central than underived /e/, thus replicating the previously observed incomplete neutralization between underived and derived /e/. This acoustic difference could not be explained as a stress, word length or orthography effect. We suggested that this example of incomplete neutralization may be the reflection at the acoustic level of different phonological representations and consequent production mechanisms. This hypothesis was tested by using an articulatory model that employs the same units both at the planning and at the execution level, and by comparing naturally-produced data to modeled stimuli. Specifically, derived /e/ was articulatorily modeled as a bi-gestural ‘blending’, reflecting its origin in the alternation with diphthong /ea/, while underived /e/ was modeled as a single gesture. The results showed that articulatorily synthesized stimuli modeled with these gestural specifications were acoustically similar to naturally produced derived and underived /e/ tokens, respectively. The stimulus modeled with two synchronously coupled gestures for /e/ and /a/ had similar acoustic properties to naturally produced derived /e/, and differed from naturally produced underived /e/. The acoustic similarity between natural and modeled stimuli was taken as an indication of a likewise similarity at the production (and planning) level, and thus the articulatory configuration probably employed in natural production could be inferred from the known articulatory configuration employed in the model.

Under the proposed analysis, the incomplete neutralization between the two types of /e/ in Romanian is assumed to be the result of their different representations and as a consequence, of their distinct production mechanisms. Derived /e/ is acoustically more centralized as a result of its bi-gestural representation/production involving a blending between a typical /e/ and an /a/. In the model, the difference between mono-gestural and 'blended' /e/ is realized in terms of constriction location (on the front-back dimension, cf. Figure 2), which translates into an acoustic difference on the F2 (front-back) dimension. The observed difference mainly in F2 between naturally produced derived and non-derived /e/ can therefore be understood as the direct result at the production level of the different representations (and associated articulations) of the two /e/ types. The short acoustic duration of derived /e/, typical for a single unstressed vowel in a multi-syllable word, further points to the plausibility that the two gestures of derived /e/ are produced synchronously, adding no extra duration, rather than sequentially.

A model that assumes transparency between the planning (representational) level and the execution (production) level, such as the one assumed here, can also account for other cases of incomplete neutralization. To briefly discuss one of the examples mentioned in the introduction, the incomplete neutralization between devoiced and underlyingly unvoiced stops can be explained as a result of their different gestural laryngeal representations: even when the timing and/or magnitude of the laryngeal setting typical for voiced consonants changes in some way, unless it changes categorically to a setting typical for voiceless consonants, the different representations should be directly reflected in production. Implications of the proposed model for other cases of incomplete neutralization, as well as supporting empirical evidence, remain however the subject of further research.

Given the assumed identity between units of representation and units of production, the model proposed here is simultaneously a linguistic model, and a model of execution, with no translation needed between the levels. As such, it allows for a direct reflection at the production level of the differences at the phonological level, meeting thus the challenge of being a linguistic model that allows the acoustic signal to be shaped by phonological differences below phonemic contrast. With respect to incomplete neutralization phenomena, it bypasses the major challenge of traditional segmental phonological models, namely that the symbolic units that serve as input for the production level bear no evidence of possible differences at the phonological/representational level. Thus, in a traditional segmental model, the segment /e/ would be produced as an /e/ regardless of whether it was underlying, or the result of a phonological process.

To account for incomplete neutralization phenomena, a segmental model would need to allow the phonetic realization of a given segment to be determined by its relation with other members of its paradigm, or by the relation between underived and derived forms (in whatever instantiation). Thus, a segmental approach must in some ad-hoc manner stipulate different representations for derived and non-derived segments, accessible at the production level, so that they end up with different phonetic realizations.

One other alternative to the proposed gestural model is to assume that representations are neither segmental nor gestural, but rather acoustic exemplars (Pierrehumbert, 2002, 2001), and that for some reason the acoustic memories of distinct classes (e.g., derived and non-derived forms), or even of different lexical items are different in exactly the manner observed experimentally. However, assuming that incomplete neutralization is maintained due to speakers' exemplar memory of the items in question has little explanatory power regarding the specific ingredients involved in production. In this sense, Exemplar Theory and the gestural model proposed here are not mutually exclusive, with the gestural model providing precisely the ingredients (gestures and timing relations) that may be involved in maintaining such memory-based detailed phonetic distinctions.

Acknowledgements

Work supported by the Deutsche Forschungsgemeinschaft (PO 1269/1-1) and by Fellowships from Yale University. I would like to thank Joaquín Romero and the PaPI 2009 audience for their very helpful comments and suggestions.

References

Benus, S., & Gafos, A. (2007) Articulatory characteristics of Hungarian 'transparent' vowels. *Journal of Phonetics*, *35*, 271-300.

Bishop, J. B. (2007) Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in /s/-aspiration. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1765--1768). Saarbrücken.

Boersma, P., & Weenink, D. (n.d.). *Praat: doing phonetics by computer*. Retrieved March 31, 2009, from http://www.praat.org/

Browman, C. P., & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology*, *6*, 151-206.

Browman, C. P., & Goldstein, L. (1992) Articulatory phonology: an overview. *Phonetica*, *49*, 155-180.

Browman, C. P., & Goldstein, L. (1990) Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*, 299-320.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984) Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America*, *75*, S22-S23.

Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996) Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170-177.

Goldstein, L., Byrd, D., & Saltzman, E. (2006) The role of vocal tract gestural action units in understanding the evolution of phonology. In *Action to Language via the Mirror Neuron System* (pp. 215-249). Cambridge: Cambridge University Press.

Hanson, H., & Stevens, K. N. (2002) A quasi-articulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLSyn. *Journal of the Acoustical Society of America*, *112*, 1158-82.

Harrington, J. (2006) An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*, 439–457.

Harrington, J., & Cassidy, S. (1999) *Techniques in speech acoustics*. Text, speech, and language technology, v. 8. Dordrecht; Boston: Kluwer Academic Publishers.

Marin, S. (2005) Complex Nuclei in Articulatory Phonology: The Case of Romanian Diphthongs. In *Selected papers of the Linguistic Symposium in Romance Languages 34th* (pp. 161-177). Amsterdam, Philadelphia: John Benjamins.

Marin, S. (2007) An articulatory modeling of Romanian diphthong alternations. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 453--456). Saarbrücken.

Nam, H., Goldstein, L., & Proctor, M. (n.d.) *TADA (TAsk Dynamics Application)*. Retrieved from http://www.haskins.yale.edu/tada_download/

Pierrehumbert, J. (2002) Word-specific phonetics. In *papers in Laboratory Phonology VII* (pp. 101-139). Berlin: Mouton De Gruyter.

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.

Port, R. (1996) The discreteness of phonetic elements and formal linguistics: Response to A. Manaster Ramer. *Journal of Phonetics*, *24*, 491-511.

Saltzman, E., & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333-382.

Schroeder, M. R., Atal, B. S., & Hall, J. (1979) Optimizing digital speech coders by exploiting masking properties of the human ear. *Journal of the Acoustical Society of America*, *66*(6), 1647-1652.

Simonet, M., Rohena-Madrazo, M., & Paz, M. (2008) Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology* (pp. 72-86). Somerville, MA: Cascadilla Proceedings Project.

Syrdal, A. K., & Gopal, H. S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, *79*, 1086–1100.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004) Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, *32*, 251-276.

Van Son, R., & Pols, L. (1990) Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, *88*, 1683-1693.

Stefania Marin
Institute of Phonetics and Speech Processing,
Ludwig-Maximilians-University Munich, Germany
marin@phonetik.uni-muenchen.de