DYNAMIC PHONETIC DETAIL IN LEXICAL REPRESENTATIONS

Christo Kirov and Adamantios Gafos

New York University and Haskins Laboratories
cnk218@nyu.edy, ag63@nyu.edu

ABSTRACT

A dynamical model of phonetic detail is presented. The model is compared to an exemplar-based model, which has been shown to offer an account of (presumed) frequency-dependent lenition processes. The dynamical model accounts for the same lenition patterns. However, there is a key difference. In contrast to the exemplar model, the dynamical model provides a handle on the timecourse of assembling phonological representations.

Keywords: exemplar theory, sound change, dynamics, timecourse

1. AN EXEMPLAR-BASED MODEL OF LENITION

In this paper, we discuss lenition mainly in the diachronic sense of change over time, rather than as the synchronic result of a phonological rule applied to an underlying representation. One diachronic example is Grimm's law, according to which Proto-Indoeuropean voiceless stops became Germanic voiceless fricatives (e.g. PIE *[t] > Gmc *[θ]). Grimm's law can be seen as part of a "weakening chain" of PIE sounds: d > t > θ > h > ∅.

The raw motivation for lenition is assumed to be a tendency to undershoot the degree of oral constriction in contexts favoring gestural weakening (e.g. in unstressed syllables, codas, or intervocalically) [1].

According to Bybee, at least some lenition processes apply variably based on word frequency [3]. One example is schwa reduction (e.g. *memory* tends to be pronounced [mɛmri]). Another is t/d-deletion (e.g. *told* tends to be pronounced [tol]). Once a lenition process has begun, Bybee's claim amounts to saying that words with high frequency will weaken more quickly over time than rare words.

Pierrehumbert has developed an exemplar-based model of phonetic detail and has shown how this model can capture four (presumed) key properties of an unfolding lenition process [8]. First, each word displays a certain amount of variability in production. Second, as embodied in Bybee's claim, the effect of word frequency on lenition rates is gradient. Third, the effect of word frequency on lenition rates should be observable within the speech of individuals; it is not an artifact of averaging data across the different generations which make up a speech community. Finally, the effect of word frequency on lenition rates should be observable both synchronically (by comparing the pronunciation of words of different frequency) and diachronically (by examining the evolution of word pronunciations over the years within each person's speech).

In Pierrehumbert's exemplar-based model, a given linguistic category is stored in a space whose axes define the parameters of the category [8]. This continuous space is quantized into discrete cells based on perceptual limits. Each cell is considered to be a bin for perceptual experiences, and Pierrehumbert considers these bins to be the actual exemplars. When the system receives an input, it places it in the appropriate bin. All bins start out empty. When an input is added to a bin, this is equivalent to the storage of an exemplar. The new exemplar is given a categorical label based on the labels of other nearby exemplars. A decay process decreases the activation of an exemplar over time, corresponding to memory decay.

During production, a particular exemplar from memory is chosen to be produced, where the likelihood of being chosen depends on how activated the exemplar is. The chosen exemplar is shifted by a bias in the direction of lenition.

Pierrehumbert's exemplar model derives the four properties of lenition discussed earlier as follows. Variability in production is directly accounted for since production is modeled as a random sampling of all the exemplars stored. Each lexical item has its own exemplars, and each production/perception loop causes the addition of a new exemplar to the set. This new exemplar is more lenited than the speaker originally intended due to biases in production, so the distribution of exemplars skews over time. In a given period of time, the number of production/perception loops an item goes through is proportional to its frequency. So, the amount of lenition associated with a given item shows gradient variation according to the item's frequency. As all processes directly described by the exemplar model occur within a single individual, lenition is clearly

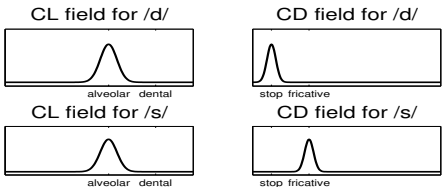

Figure 1: Component fields of /d/, /s/. y-axis represents activation. /d/ and /s/ have nearly identical CL fields, as they are both alveolars, but they differ in CD.

observable within the speech of individuals. Diachronically, lenition will proceed at a faster rate for more frequent items because they go through more production/perception loops in a given time frame. The synchronic consequence of this is that at a point in time, more frequent items will be more lenited in the speech of an individual than less frequent items.

In sum, the exemplar-based model offers a direct way to represent the fine phonetic substance of linguistic categories. The model also offers a way to capture the assumed effects of frequency on the unfolding lenition process.

2. A DYNAMICAL MODEL

Although our model does not commit to any specific phonological framework or model, we will follow Articulatory Phonology [2] in order to clarify concepts. Thus, let us assume that lexical items take the form of gestural scores (we could have said segments with features or orosensory variables [5]). A gestural score, for current purposes, is simply a sequence of gestures (we put aside the intergestural temporal relations that also must be specified as part of a full gestural score). For example, the sequence /das/ consists of three oral gestures - a tongue tip gesture for /d/, a tongue dorsum gesture for /a/, and a tongue tip gesture for /s/. Gestures are specified by target values for the *vocal tract variables* of constriction location (CL) and constriction degree (CD), parameters defining the target vocal tract state. For example, /d/ and /s/ have the CL target value [alveolar], an actual numerical value in the model of [2]. The CD value of /d/ is [stop] and for /s/ it is [fricative].

The first crucial part of our proposal concerns the way in which these parameters are specified. In our model, each parameter is not specified by a unique numerical value as above, but rather by a continuous activation field over a range of values for the parameter. Fields then resemble distributions depicting the continuous details of vocal tract variables. A lexical item therefore is a gestural score where the parameters of each gesture are represented by their own fields. Schematic fields corresponding to the (oral) gestures of the consonants in /das/ are given in Figure 1.

In our model, lexical items are manipulated using the formalism of Dynamic Field Theory [4], henceforth DFT. In a DFT formalism, gestural parameters are represented using continuous activation fields. The basic dynamics governing each field are described by the following equation:

$$\tau \frac{dp(x,t)}{dt} = -p(x,t) + h + s(x,t) \quad (1)$$

where p is the field in memory (a function of a continuous variable x), h is the field's resting activation level (a lower limit on activation), τ is a constant corresponding to the rate of decay of the field (i.e. the rate of memory decay), and s is a field representing time dependent external input to the system (i.e. perceived token) in the form of a localized activation spike.

The $\tau dp(x,t)/dt = -p(x,t)+h$ part of the equation is fundamentally the same as the exponential decay equation, $dx/dt = -x$, but each trajectory is shifted on the y-axis by h. Much as the parameter x decays in the exponential equation, the activation along the field p wanes over time and falls to its resting level h, unless the other terms in the equation slow or reverse the process. The input term $s(x,t)$ represents sensory input. More specifically, input is a peak of activation registered by the perception module. This peak is located at some detected x-axis value along the field. This value is assumed to be sub-phonemic in character. For example, we assume that speakers can perceive gradient differences in Voice Onset Time values, constriction location, and constriction degree within the same phonemic categories. In the current model, the input term is formulated as $e^{-(x-off)^2}$, where off is the detected value, or offset, along the x-axis of the field.

Once input is presented, a field can evolve to a stable attractor state, that is, a localized peak at a value corresponding to the input. The state is stable in the sense that it can persist even after the input has been removed. In effect, the field for the lexical item has retained a memory of the sub-phonemic detail in the recently perceived input (a concrete example is presented later in Figure 4).

We now turn to the second and perhaps most crucial part of our proposal. Formalizing lexical representations with dynamic fields adds a timecourse dimension to the gestural specification process. If a lexical representation contains a /d/, the CD and CL parameters for this /d/ are not statically assigned to their (language- or speaker- specific) canonical

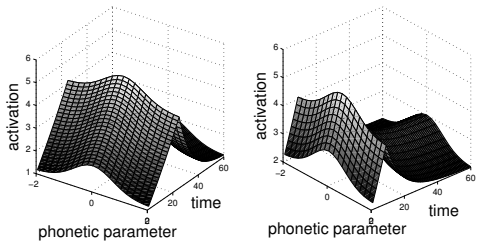

Figure 2: Sampling was simulated with a decision threshold $d = 5$ and $\tau = 10$. The first field (left, $h_0 = 1$) reached the decision threshold at $t = 25$, and the second field (right, $h_0 = 2$) reached the decision threshold at $t = 9$ (where t is an arbitrary unit of simulation time). The field with higher initial resting activation reached the decision threshold faster.

values, e.g., CL = [alveolar]. Rather, assigning values to these parameters is a time-dependent process, captured as the evolution of a dynamical system over time. In short, lexical representations are not static units. This allows us to derive predictions about the timecourse of assembling different lexical representations.

The specification process begins by a temporary increase in the resting activation of the field (i.e. pushing the field up) caused by an intent to produce a particular lexical item. Resting activation increases steadily but noisily until some part of the field crosses a decision threshold and becomes the parameter value used in production. This scheme ensures that the areas of maximum activation are likely to cross the decision threshold first. After a decision has been made, resting activation returns to its pre-production level.

The gestural specification process is affected by the pre-production resting activation of the field, in that a field with high resting activation is already "presampled", and thus automatically closer to the decision threshold. This leads to faster decisions made during production of more activated fields, and by extension more frequent lexical items. We have confirmed this behavior by computer simulation. See Figure 2 which compares two fields identical in all respects except for resting activation. The field with the higher resting activation reaches the decision threshold first. This prediction has been confirmed experimentally by studies showing that pictures of more frequent words are named faster than pictures of less frequent words [6].

In the exemplar model, neither production nor perception has any timecourse. Both processes are instantaneous. Thus, the effect of frequency in the speed of producing different lexical items cannot be accounted for in that model. See [7] for further exemplification of timecourse predictions.

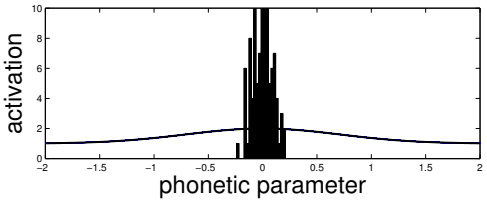

Figure 3: Variability in production. Histogram of selected values over 100 simulations of gestural specification. Histogram overlaid on top of field to show clustering of selected values near the field maximum.

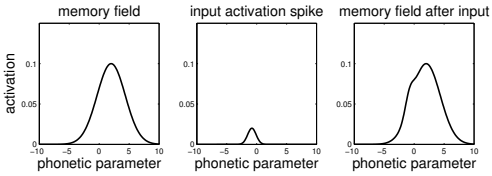

Figure 4: (Left) Field of lexical item in memory. (Middle) Input function (output of perception corresponding to $s(x, t)$ in equation 1). (Right) Field of lexical item in memory after input is added to it. Field shows increased activation around area of input.

In addition, the noisy character of the specification process allows for variation in the value ultimately specified. Figure 3 shows the results of a series of simulations verifying that although there is variation in selected values, they cluster reliably around the maximally activated point of the field.

At present, we only model how parameter fields in the lexicon are affected during speech recognition. Activation in each lexical field increases slightly around the range of the input parameter value. More precisely, the perceptual system converts a heard production into an input function $s(x, t)$. This spike is directly added to the appropriate field, resulting in increased activation at some point along the field's x-axis (see Figure 4). Since activation fades slowly over time, only areas of the field that receive reinforcement are likely to remain activated. So, a peak in activation may shift over time depending on which region of the field is reinforced by input.

2.1. Lenition in the dynamical model

Given an initial field representing the current memory state of a lexical item, we can simulate lenition using the dynamical model described above. Figure 5 shows the results of one set of simulations. Shown are the state of the simulation at the starting state, after 50 samples of a token, and after 100 samples. Each time step of the simulation corre-

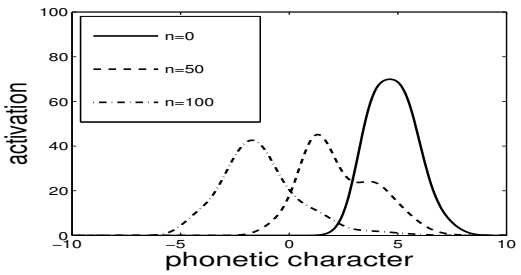

Figure 5: Output of lenition simulation. x-axis represents a phonetic dimension (e.g. constriction degree during t/d production). Each curve represents a distribution of a particular category over the x-axis at a point in time. As time progresses, the distribution shifts to the left (i.e. there is more undershoot/lenition) and becomes broader.

sponds to a production/perception loop. Production was performed as described above by picking a value from the field and adding noise and a bias to it. This produced value, encoded by an activation spike of the form $e^{-(x-off)^2}$, where $off = sample(p) + noise + bias$, was fed back into the system as input.

As can be seen in Figure 5, at the point when lenition begins, the field represents a narrow distribution of activation and there is little variability when sampling the field during production. As lenition progresses, the distribution shifts to the left. During this time the distribution becomes asymmetrical, with a tail on the right corresponding to residual traces of old values for the parameter. It also grows wider, corresponding to an increase in parameter variation while the change occurs.

Like the exemplar model above, the dynamical model can derive the four lenition properties. Variability in production is accounted for by noise during the gestural specification process. Each lexical item has its own fields and each production/perception loop causes a shift in the appropriate field towards lenition due to biases in production (see Figure 4 for an example of a field starting to skew to the left). In a given period of time, the number of production/perception loops an item goes through is proportional to its frequency. So, the amount of lenition associated with a given item shows gradient variation according to the item's frequency. All the processes described here occur within a single individual, so lenition is clearly observable within the speech of individuals. Diachronically, lenition will proceed at a faster rate for more frequent items, again because they go through more production/perception loops in a given time frame. This same mechanism is evident synchronically as well, since at any single point in time, more frequent items will be more lenited than less frequent items.

3. CONCLUSION

The dynamical model of phonetic detail presented here remains consistent with one key aspect of generative theories of representation. Instead of representing categories extensionally as arbitrarily large exemplar sets, linguistic units and their parameters can have singular representations. These are the dynamic fields in our specific proposal. It is these unitary representations, rather than a token by token expansion of the exemplar sets, that drifts in sound change. Second, the dynamical model inherently deals with time. Since both the exemplar and the dynamical model are at least programmatically designed to include production and perception, which unfold in time, this seems to be a key property.

4. ACKNOWLEDGMENTS

Research supported by NIH Grant HD-01994 to Haskins Labs. AG also acknowledges support from an Alexander von Humboldt Research Fellowship.

5. REFERENCES

[1] Beckman, M. E., de Jong, K., Jun, S.-A., Lee, S.-h. 1992. The interaction of coarticulation and prosody in sound change. *Language and Speech* 35, 45–58.

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

[3] Bybee, J. 2003. Lexical diffusion in regular sound change. In: Restle, D., Zaefferer, D., (eds), *Sounds and Systems: Studies in Structure and Change*. Mouton de Gruyter, Berlin 58–74.

[4] Erlhagen, W., Schöner, G. 2002. Dynamic field theory of movement preparation. *Psychological Review* 109, 545–572.

[5] Guenther, F. H. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102, 594–621.

[6] Jescheniak, J., Levelt, W. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological forms. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 824–843.

[7] Kirov, C., Gafos, A. On the timecourse of assembling phonological representations. In: Chitoran, I., Coupè, C., Marsico, E., Pellegrino, F., (eds), *Approaches to Phonological Complexity*. Mouton de Gruyter. To appear.

[8] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J., Hopper, P., (eds), *Frequency Effects and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam 137–157.