

Shifting Categories: An Exemplar-based Computational Model of Chain Shifts

Marc Ettliger (marce@berkeley.edu)

Department of Linguistics, 1203 Dwinelle Hall
Berkeley, CA 94720 USA

Abstract

Explanations for historical chain shifts tend towards the teleological, using abstract ideas like *balance* and *equilibrium* as the organizing principles of a language's sounds. This paper investigates whether there are more basic phonetic principles governing the behavior of sound categories with respect to one another. Using a computational simulation of agents communicating with each other, I show that vowel chain shifts fall out naturally from an exemplar-based model of sounds. This suggests that no overarching teleological mechanisms are required to account for chain shifts and that the self-organizing behavior of exemplar-based categories provides an adequate explanation.

Keywords: computational modeling; linguistics; exemplars; categorization; communication.

Introduction

Historical chain shifts are a type of sound change where the change of one sound triggers the change of another. For example, the Great Vowel Shift (Jespersen, 1949) consisted of a chain shift of the long vowels in English. In one part of the change, the first formant of /e:/, (as in *let*), decreased so as to sound like /i:/ (*leek*). Accompanying this change was a decrease in the first formant of /a:/, (*lack*), to an /e:/, (*lake*). Vowels can be conceptualized as being situated in a two dimensional space according to their first and second formant frequencies so a representation of the Great Vowel Shift can be illustrated as in figure 1:¹

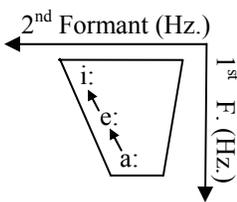


Figure 1. A portion of the Great Vowel Shift

The interconnectedness of these changes and those in other similar shifts have been used as evidence that the overall structure of the sounds of a language are regulated. However, the explanations for this regularization are often teleological such as the assertion that a disturbance in the balance of a system must be restored (Jakobson, 1931) or that “the system exerts a pull on sounds that aren’t fully integrated” Martinet (1955). Even contemporary authors have described chain shifts as sounds “moving to *fill the slot*

¹ Placing the origin at the top right allows the vowel space to approximate tongue position for the articulation of the vowels.

vacated by the movement of another” (Hock and Joseph, 1996) [emphasis added].

This begs the question of whether the movement of vowels is indeed regulated by abstract forces like balance and the pull of a system or whether there are more basic cognitive principles at work. The hypothesis explored here is that it is the nature of exemplar-based categories themselves that provides the basis for the behavior exhibited by vowel systems.

Using a computational simulation of language-using agents, I show that an exemplar-based model of the production and perception of vowels predicts the shift of one vowel in the same direction as the shift of another (as in the graphical representation in figure 1) without resorting to external constraints. This demonstrates that, in certain respects, phonological categories are *self*-regulating and that the appearance of a balanced system is epiphenomenal of more basic properties of exemplar categories.

The basic schema of the interconnected change is as follows: In the case of the /i/-/e/ chain shift, this behavior arises because as /i/ begins to move further away from /e/ it also shifts the boundary between the two. Therefore, vowel tokens that once would have been ambiguous between /i/ and /e/ or just barely categorized as /i/ now fall within the boundary of /e/ and are categorized as such. These tokens become part of the /e/ category and begin to shift /e/ towards the former boundary. As /i/ continues to move further and further away from its original location, /e/ continues to move towards it, until /e/ occupies /i/’s former place in the vowel space.²

Maximizing the Vowel Space

The linked sound changes that make up a chain shift can be construed as being the result of maximizing the vowel space and the maximally dispersed vowel space found cross-linguistically (Maddieson, 1984) has been modeled in a number of ways. Liljencrants and Lindblom (1972) and Schwartz, Boë, Vallée and Abry (1997) both used the idea of maximizing acoustic distinctiveness to account for the most typical vowel inventories found cross-linguistically. There are two limitations of an account of this sort. First, there are a number of languages that simply do not maximize the vowel space, including Abkhaz (ə, e, a; Trubetzkoy, 1930), Nahuatl (i, e, o, a; Andrews, 1975) and Manobo (i, ī, ə, a, o; Meiklejohn and Meiklejohn, 1958) to

² This is not an account of the Great Vowel Shift, which is not believed to have begun with a decrease in the F1 of /i/. The use of the GVS and figure 1 simply serves as a well-known example of the basic principles of a chain shift.

name but a few. Second, these accounts describe a synchronic state of the inventory without explaining *how* the sounds become optimized. A more satisfactory explanation is one that explains the mechanisms by which these patterns emerge.

A number of recent proposals have attempted to do just that. Joanisse and Seidenberg (1997) used a neural network to show that a maximally dispersed vowel space is easier to learn, but their simulation only looked at five-vowel systems. De Boer (2000) proposed a model based on communicating agents which involved imitation, feedback, recognition of intention and success rate tracking. This mode of interaction added to the complexity of the simulation and do not have clear correlates in real human communication. Oudeyer (2005) eliminated this complexity by coupling the motor and perceptual maps through a neural network to achieve similar results.

The model presented here differs from these previous accounts in two crucial respects. First, it does not require complex interaction between agents and avoids the assumptions implicit when using neural networks. Instead, it simply posits that cognitive categories exhibit properties of exemplar-based categories. Therefore, the results only reflect the self-organizing behavior of exemplar-based categories and not explicit linguistic capabilities of speakers.

Second, all of the previous models explain the well-attested balanced vowel systems, but do not have an explanation for the un-balanced or non-optimal systems mentioned above. This is the result of focusing on constraints on the final state of the vowel space instead of focusing on the diachronic (historical) processes of *change* leading to that final state. This model differs in that it seeks to model the actual process by which vowels change.

A Model of Exemplar-based Agents

The simulation presented here models the diachronically observed phenomenon of a chain shift by simulating agents talking to each over some span of time. Communication consists of the production of a random utterance by a random agent (“speaker”) that is perceived and categorized by another (“hearer”). The representation of the agent’s vowels and the processes of producing and perceiving these vowels are governed by exemplar theory.

First introduced in psychology as a theory of categorization (Medin and Schaffer, 1978), exemplar theory’s application to linguistics (Johnson and Mullenix, 1997; Pierrehumbert, 2001; Wedel, 2004) is based on the idea that each of the sounds of a language corresponds to a category which is represented by a cloud of the remembered tokens of that category. The next section presents a framework of how such a theory can be modeled computationally and is partially based on the model presented by Pierrehumbert (2001).

Vowel Categories

Exemplar-models are motivated for use in Phonetics by experimental evidence that has shown that significant

amounts of detailed information about heard speech are retained by listeners (Johnson and Mullenix, 1997). So, the category for the vowel /i/ might consist of a token representing the most recent time it was heard, perhaps on a roller-coaster as part of the word *Wheeeeeee!*, and another token for when it was heard in a lecture four years ago as part of the word *theory* and so on.

This does raise the problem that over the life span of a speaker, the sheer number of exemplars would grow intractably large for both brain and computer. This is resolved in two ways (Johnson and Mullenix, 1997; Pierrehumbert 2001).

First, the memories decay as time progresses so that more recent members of a category have greater strength. This corroborates experiments showing the declining recollection of specific exemplars over time (Goldinger, 1996). Decay is achieved by decrementing the strength of each exemplar exponentially over time according to the equation:

$$(1) S_t = S_0 \cdot e^{(-t/\tau)},$$

where S_t is weight at time t , τ is the time decay constant and S_0 is the initial exemplar strength. Eventually, an exemplar’s strength decreases to where it has no effect on production or perception and is dropped from the exemplar cloud.

Second, implementation involves parameters for the exemplars that are discrete, rather than gradient.

The use of discrete, rather than gradient, parameterization is based on evidence that listeners can not distinguish between stimuli that differ below a certain threshold or Just Noticeable Different (JND; Kruschke, 1992). The first simplification made is that the parameters that distinguish one vowel from the next are primarily the first and second formant frequencies and that only the first formant (F1) is involved in the chain shift (see figure 2 for a comparison of /i/ and /e/). While Carlson, Fant, and Granström (1975) and Zahorian and Jagharghi (1993) show that other formants and cues affect vowel categorization, only F1 is used for the exemplars here for purposes of simplifying exposition. Future simulations will incorporate all relevant formant frequencies and cues such as length and formant transitions.

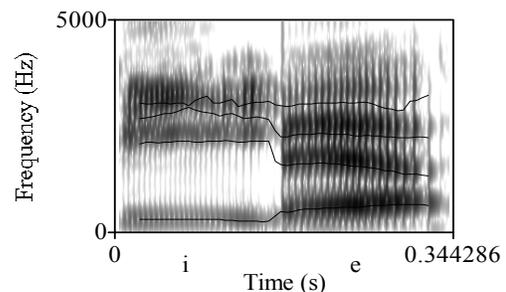


Figure 2: Formants plot of /i/ & /e/

The JND for frequency is approximately 0.5% (Roederer, 1973). So, an exemplar for the vowel /i/ with an F1 of 285

Hz would be stored along with one having an F1 of 286 Hz, which is less than .5% different. JND therefore allows the gradient frequency value to be associated with a frequency value quantized into units of $200 \times 0.5\% = 1$ Hz with 200 Hz being the lowest formant frequency of any vowel (Boë, et. al. 1989).

The model thus far consists of a stored set of discrete exemplars, each with a decaying strength and a formant frequency representing the agents' vowels.

Production

Because the focus of this model is on the behavior of vowels, each act of communication is simplified to the speaker saying only vowels to the hearer. Further simplifications are that all vowels appear with equal probability and that each agent has an equal probability of being a speaker or hearer. In reality, vowels occur with different regularity, which potentially affects their pronunciation and in turn, their formant frequency (Bybee, 2001). This can potentially account for why sometimes a chain shift occurs and other times, the two vowels merge. Also, a real network of interactions is more complex and can give priority to the pronunciations of certain speakers over others (Milroy and Milroy, 1992).

The production algorithm for a given vowel involves the activation of a random exemplar within the cloud making up the vowel category, weighted by exemplar strength. The formant values of vowels for an individual speaker approximates a normal distribution (Wells, 1962), so variance with a standard deviation of 45 Hz (Wells, 1962) is added to the selected exemplar frequency and the corresponding formant frequency is then "said" to the hearer.

Perception

The listener categorizes this new vowel token in one of its existing categories by looking at the nearby exemplars and assigning the new token to the category with the greatest exemplar strength. For example, in figure 3, a token with an F1 of 350 Hz is heard and the nearest exemplars for the two closest vowels are shown. The figure shows more /e/ exemplars near the new 350 Hz token than there are /i/ exemplars and the /e/ exemplars are stronger, so the new token is categorized as an /e/ and added to the cloud of exemplars for the /e/ category.

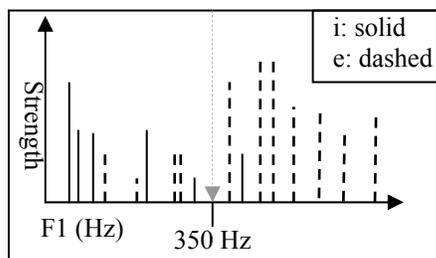


Figure 3. Addition of a new exemplar

The equation for calculating the activation of each category is based on a fixed window of activation (Pierrehumbert, 2001; Lacerda, *in press*) and is calculated as:

$$(2) \text{ activation}(x) = \sum_{i=x-w \dots x+w} S_i$$

where x is the token's frequency, w is the frequency window and s_i is the exemplar strength of exemplars at i frequency. The impact of different window sizes is discussed in the Parameters section while the same results obtain if the effect of all exemplars are used, weighted in inverse proportion to their distance from the new exemplar (Kruschke, 1992). The new exemplar is added to the exemplar cloud of the category with the greatest activation.

Aging + Cycles

After a parameterized number of token exchanges, all of the agents' exemplars are aged according to the equation in (1). This production-perception-aging sequence is repeated for a parameterized number of cycles. Vowel chain shifts are reported to occur with the course of a couple of generations (Gordon, 2001), so the number of cycles is set to approximate the number of times a person hears a vowel over the course of a lifetime. Approximations are on the order of 10,000 tokens/month, so the number of cycles is ~10,000,000. The results below, however, show effects after 10,000 exchanges.

Modeling Chain Shifts

This section describes the successful implementation of the above model in simulating a chain shift whereby changing the quality of one vowel changes the quality of another. For the sake of brevity only two vowels are used, /i/ and /e/, but the result can be extrapolated to any number of vowels since the principles stay the same. In the model, /e/ shifts as /i/ shifts, and by the same mechanism, /a/ also shifts as /e/ shifts.

Initial State

The description of a chain shift begins with a vowel system that is in some sort of temporary equilibrium, as was the case with English vowels some 500 years ago. This is represented by the initial state, or seed values, of the computational model. A temporary equilibrium can be represented by a single speaker with a single F1 exemplar for each of the vowels (/i/ and /e/). Since the model seeks to investigate whether an incremental decrease in F1 for one vowel automatically decreases the F1 of the other vowel something needs to trigger the decrease in F1 of one of the vowels.

There are a number of possible acoustic, physiological and sociolinguistic factors that may contribute to the changing formants of a vowel (Ohala, 1993). The aim is not to model those factors, but rather to assume that one vowel changes and model why another vowel in the same system

must necessarily change. If any one vowel has a 90% chance of changing due to the above factors, we should expect that the chance that two vowels change is 81%, three vowels 73% etc. Instead, however, the historical pattern shows that the vowels move in concert.

The initial formant frequency decrease in the model is set into motion by introducing another speaker with a lower F1 than the first for /i/. The initial change could also have been initiated by including an articulatory bias in one of the vowels of one speaker. To ensure that any change found in /e/ is based on the change in /i/ and not on the /e/ of the second speaker, the initial F1 for /e/ is kept the same. This simulates contact with another language or community of speakers with different vowel qualities, which is one of the known triggers for this type of sound change (Labov, 2001). This is shown in table 1:

Table 1. F1 formant frequencies of agents' vowels at initial state (in Hz):

Vowel	Agent 1	Agent 2
/i/	285	215
/e/	570	570

Results

If only the /i/ vowel is considered, then the expectation is that after enough cycles, the F1 of the two agents will eventually converge and average to somewhere around $(285+215)/2 = 250$ Hz since the strength of each exemplar is the same and each agent has an equally likely chance of being a speaker or hearer. If the exemplar strength for the /i/ of one of the agents was higher than the other, we would expect that the eventual F1 that the two agents converge upon would be somewhere closer to the agent with the higher strength.

The results showing the average F1 of /i/ for both agents after n exchanges for 500 trials of the model are shown in table 2. After ~10,000 exchanges, the vowels converge to around 250 Hz achieving the desired gradual decrement of F1 for agent 1's /i/.

Table 2. Convergence of F1 for /i/ (in Hz).

Agent	Exchanges				
	0	10	100	1000	10,000
Agent 1	285	283	254	247	247
Agent 2	215	221	226	233	242

Table 3. F1 for /i/ and /e/ for both agents (in Hz).

Agent	Vowel	Exchanges				
		0	10	100	1000	10,000
Agent 1	/i/	285	281	273	261	244
	/e/	570	568	557	540	524
Agent 2	/i/	215	216	227	238	245
	/e/	570	571	566	550	541

The data in table 3 show the average of 100 new runs for both vowels over the course of 10,000 exchanges and is also graphed in figure 4.

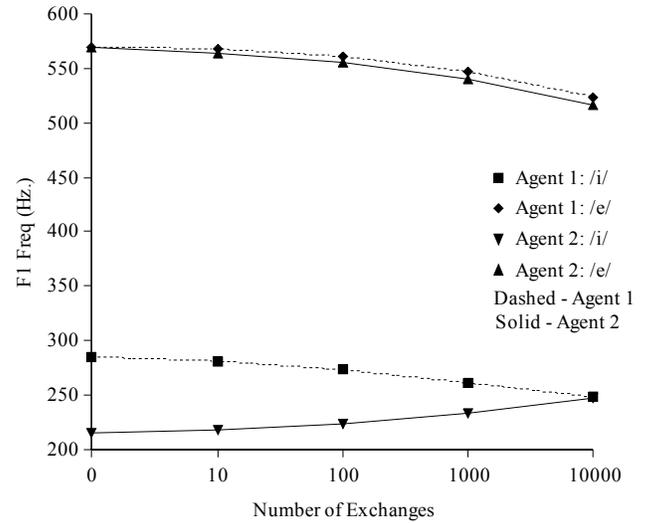


Figure 4. F1 formant frequency for all vowels

The key number to look at is the formant frequency of the vowel /e/ for agent 1. As the table and graph show, there is an appreciable decrease in this value from the initial seed (570 Hz) to the final average after 10,000 exchanges (524 Hz). This decrease can not be due to influence of the F1 frequency of the second agent because that agent starts with the same frequency for /e/. Instead it is the decrease in F1 of the vowel /i/ that is causing the decrease in the F1 of /e/. This is precisely the behavior of chain shifts that this simulation attempts to model.

Parameters

Most of the arbitrarily set parameters affect only the rate or degree at which the /e/ shifts and not the existence of a shift in and of itself. Repeated runs of the model with different parameters show that the initial strength of the exemplars of the agents, the ratio of cycles to exchanges and the number of agents also do not change whether the shift occurs.

However, the relationship of the variance of the two vowels to the perception window does play a significant role in determining whether the chain shift happens or not. When the sum of the two parameters drops below ~0.5 the distance between the two vowels in question, the chain shift ceases to occur. This suggests that chain-shifts are closely tied to the amount of variation in the articulation of vowels: The greater the variance, the more likely a chain shift is to occur. This results because as the variance and/or perception window decreases, the chance a token falls somewhere near the boundary between the two categories becomes vanishingly small.

Conclusion

These simulations show that the nature of exemplar-based categories is such that when one vowel moves, another will fill the gap. The vowel space, therefore, is *self-organizing* in that its harmony and balance are maintained simply by virtue of how exemplar categories operate.

An examination of the behavior of each exchange suggests that the mechanism by which this happens is as described above. The process begins when the center of gravity of the exemplar cloud for /i/ begins to shift downward (in Hz). This shifts the boundary – the frequency where the value for the equation in (2) is the same for both vowels – down as well. Once this happens, more and more tokens near that boundary begin to be categorized as /e/. The categorization of these low-F1 tokens as /e/ exemplars shifts /e/'s center of gravity, and therefore the boundary, down even further. As /i/ continues to change, so does /e/, and the well attested historical chain shift is observed.

This model also has the advantage of providing an explanation for why chain shifts sometimes happen and sometimes do not: vowels with a lower degree of variance do not engage in chain shifts.

There are myriad factors that interact with the exemplar-based self-organization modeled here including speech production and perception constraints (Schwartz et. al., 1997), the effect of other phonological units, the influence of phonemic contrast and social influences that all impact the ultimate behavior of the categories. These other factors would better account for why long vowels tend to shift upwards in the vowel space while short vowels lower (Sweet, 1888), for example. The current model also over-predicts the types of sound change found in that it would also predict an unattested chain shift for F2, or a downward chain shift for long vowels. This account should therefore be construed as being a piece of the larger puzzle that chain-shifts presents and not a complete account.

Finally, the model also has potentially significant ramifications for any category system that is parameterized along a scalar dimension similar to the F1 formant frequency used here. In addition to the attested repercussions of the perceptual magnet effect (Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992) this model suggests that there may be other emergent properties of exemplar-based categories.

Further modeling will hopefully capture the interaction of vowels in multiple dimensions and potentially provide insight into consonant chain shifts, why most, but not all, languages have maximally dispersed vowels and how the intricacies of social interaction affects these phonological properties.

Acknowledgments

I would like to thank Andrew Garrett for the initial inspiration to investigate these questions, Keith Johnson for continuing feedback throughout the project and the anonymous reviewers whose comments proved invaluable.

References

- Andrews, J., & Richard, D. (1975). *Introduction to classical Nahuatl*. Austin: Univ of Texas Press.
- Boë, L., Perrier, P., Guérin, B., & Schwartz, J. (1989). Maximal vowel space, *EuroSpeech 1989*, 2: 281-284.
- Carlson, R., Fant, G., & Granström, B. (1975) Two-formant models, pitch, and vowel perception. In Fant, G. & Tatham, M. (eds.) *Auditory Analysis and Perception of Speech*. 55-82, Academic Press, London.
- de Boer, B. (2000). Self organization in vowel systems, *Journal of Phonetics* 28, 441–465.
- Flemming, E. (1995). *Auditory features in Phonology*. UCLA Ph.D. Dissertation.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Gordon, M. J. (2001). *Small-town Values and Big-city Vowels: A Study of the Northern Cities Shift in Michigan*. Durham, N.C.: Duke University Press.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition* 22: 1166-1182.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411–428.
- Hock, H. & Joseph, B. (1996). *Language History, Language Change and Language Relationship*. New York: Mouton de Gruyter.
- Jakobson, R. (1931) Über die phonologischen Sprachbünde. *Travaux du Cercle Linguistique de Prague* 4.
- Joanisse, M., & Seidenberg, M. (1997). [i e a u] and sometimes [o]: Perceptual and computational constraints on vowel inventories. *Proc. of the 15th Conf. of the Cog. Sci. Soc.* Stanford: CA.
- Johnson, K. & Mullenix, J. (1997). (eds.). *Talker Variability in Speech Processing*. San Diego, California: Academic Press.
- Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psych. Rev.* 99, 22-44.
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 255, 606-608.
- Labov, W. (1994). *Principles of linguistic change. Volume I: Internal Factors*. Oxford: Blackwell.
- Labov, W. (2001). *Principles of Linguistic Change. Volume 2: Social factors*. Oxford: Blackwell.

- Lacerda, F. *in press*. Distributed memory representations generate the perceptual-magnet effect. *Journal of the Acoustical Society of America*.
- Liljencrants, J., Lindblom B. (1972). Numerical simulation of vowel quality. systems: The role of perceptual contrast. *Language* 48, 839-862.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Martinet, A. (1955). *Economie des changements phonétiques*. Berne: Francke.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Milroy, L. & Milroy, J. (1992). Social network and social class: Toward and integrated sociolinguistic model. *Language in Society*. 21:12.
- Meiklejohn, P., & Meiklejohn, K. (1958). Accentuation in Sarangani Manobo. *Oceania Linguistic Monographs* 3, 1-5. Sydney: University of Sydney.
- Ohala, J (1993). The phonetics of sound change. In Jones, C. (Ed.) *Historical Linguistics: Problems and Perspectives*. London: Longman. 237-278.
- 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.
- Roederer, Juan. (1973). Introduction to the Physics and Psychophysics of Music. The English Universities Press: London.
- Schwartz, J., Boë, L., Valle, N., & Abry, C., (1997). Major trends in vowel systems inventories. *J. Phonetics* 25.
- Sweet, H. (1888). *The History of English Sounds*. Oxford, Clarendon Press.
- Trubetzkoy, N. (1930). Nordkaukasische Wortgleichungen. Wiener Zeitschrift für die Kunde des Morgenlandes. Bd. XXXVII, Heft 2. Wien.
- Wedel, Andrew. (2004). Self-organization and categorical behavior in phonology. PhD. Dissertation. UC Santa Cruz.
- Wells, J.C., (1962). *A study of the formants of the pure vowels of British English*. Unpublished MA thesis, University of London.
- Zahorian, S. & Jagharghi, A. (1993) Spectral-shape features versus formants as acoustic correlates for vowels. *J. Acoustical Soc. Am.*, 94, 1966-1982.