

A rational account of perceptual compensation for coarticulation

1. Introduction

- *Perceptual compensation* (PC): Depending on their acoustic, phonological, semantic, and syntactic contexts, two identical acoustic segments can be labeled differently and two different acoustic segments can be labeled as identical.
- Examples:
 1. Beddor et al. (2002): Adult English, Shona speakers asked to categorize V_1 from an /a/~e/ continuum (as /a/ or /e/) in pV_1pV_2 contexts ($V_2=/a/$ or /i/).
 - /e/ heard more often in VC/a/ context
 - /a/ heard more often in VC/i/ context
 - \implies speakers compensate for vowel-vowel coarticulation effect on F1.
 - Size, type of compensation effects language-specific.
 - Modified replication of English experiment below (Sec. 4).
 2. Lotto et al. (1997): Japanese quail habituated to aC_1C_2a stimuli ($C_1=/l/, /r/, C_2=/d/, /g/$), pos. reinforc. for peck at final /ga/, then peck rate measured for C from /d/~g/ continuum in alCa, arCa contexts.
 - Similar human task (pecks \rightarrow 2AFC) (Lotto and Kluender, 1998)
 - Human and quail results similar: /g/ heard more often in al context, /d/ more often in ar context.
 - \implies PC not human-specific, need general (non-human) account.
 3. PC very general: Four-month-olds (Fowler et al., 1990), non-native phonemes (Mann, 1986), nasalization (Beddor and Krakow, 1999), non-speech stimuli (Lotto, 2004), ...
- **Linguistic questions:**
 1. How can a listener misperception-based model of sound change (Ohala, 1993) be reconciled with robustness of PC?
 2. Must hypocorrective sound change result only when the conditioning context is not detected by the listener?
- Today:
 - Simple probabilistic model that explains PC as a consequence of optimally categorizing speech sounds in the presence of contextual variation.
 - Listeners' perception of speech sounds, which is biased toward the means of phonetic categories (Feldman and Griffiths, 2007), is conditioned by phonetic variation introduced by the context, as well as category frequencies in that context.

2. Previous work

- Proposed explanations for PC:
 1. Lexical, phonemic content of context (Elman and McClelland, 1988; Samuel and Pitt, 2003)
 2. Gestural theorists (Direct Realist) (Fowler, 1996, 2006)
 3. General auditory processes (frequency contrast, spectral contrast) (Kingston, 1992; Kingston and Diehl, 1995; Lotto and Kluender, 1998)
- None general enough to provide unified account of PC effects.

3. Model

- Following Feldman and Griffiths (2007) analysis of perceptual magnet effects.
- *Rational analysis* (RA) (Marr, 1982; Anderson, 1990): Attempts to explain aspects of cognition as adaptive responses to the environment.
- Basic assumption/hypothesis: People behave optimally when performing cognitive tasks.
- “Optimal” can be understood in terms of Bayesian inference (e.g. Tenenbaum and Griffiths, 2001), maximum likelihood (e.g. Fried and Holyoak, 1984), etc.; will use Bayesian inference here.
- Assume:¹
 1. Listeners hear signal S in context k , must decide its probability of belonging to category c_1 or c_2 .
 2. S normally-distributed around a target pronunciation T , itself normally-distributed around a category mean.

- Formally,

$$T | c_i, k \sim N(\mu_{c_i, k}, \sigma_c), \quad S | T, c_i \sim N(T, \sigma_S)$$

where:

- $\mu_{c_i, k}$: Category i mean in context k
- σ_C^2 : Variance in T around category mean.
- σ_S^2 : Variance in S around T .
- Note: assuming for simplicity that σ_C, σ_S same for categories 1 and 2.
- The probability of a listener finding S to have come from category c_1 can be calculated with Bayes’ rule:

$$P(c_1 | S, k) = \frac{P(c_1 | k)P(S | c_1, k)}{P(c_2 | k)P(S | c_2, k) + P(c_1 | k)P(S | c_1, k)} \quad (1)$$

- $P(c_i | k)$ is the probability of category i occurring in context k , i.e. in the lexicon as a whole.
- The $P(S | c_i, k)$ are calculated by integrating over all possible T . Eventually (derivation in App. A) we get

$$P(c_1 | S, k) = \left(1 + \frac{f_2}{f_1} e^{b-Sg}\right)^{-1} \quad (2)$$

where

$$b = \frac{1}{2} \frac{\mu_{c_1, k}^2 - \mu_{c_2, k}^2}{\sigma_S^2 + \sigma_C^2}, \quad g = \frac{\mu_{c_1, k} - \mu_{c_2, k}}{\sigma_C^2 + \sigma_S^2}$$

and $f_1 = P(c_1 | k)$, $f_2 = P(c_2 | k)$ (category frequencies).

- Eqn. 2 is a logistic function with bias b and gain g .
- Setting $P(c_1 | S, k) = 0.5$ in (2) gives the *crossover point* S' :

$$S' = \frac{\mu_{c_1, k} + \mu_{c_2, k}}{2} + \frac{\sigma_S^2 + \sigma_C^2}{\mu_{c_1, k} - \mu_{c_2, k}} \ln\left(\frac{f_2}{f_1}\right) \quad (3)$$

- Note that if it is assumed that $f_1 = f_2$, S' is simply halfway between the category means.

4. Example

- To illustrate more concretely, we show predicted and observed identification curves for PC for vowel-to-vowel coarticulation in English.

¹We assume here that compensation is in T , i.e. that T is the variable shifting by context. If we instead assume compensation is in S , of the form

$$T | c_i, k \sim N(\mu_{c_i}, \sigma_c), \quad S | T, c_i, k \sim N(T + \Delta_{i, k}, \sigma_S)$$

all results turn out the same. It thus does not matter under this analysis whether compensation is in S or T .

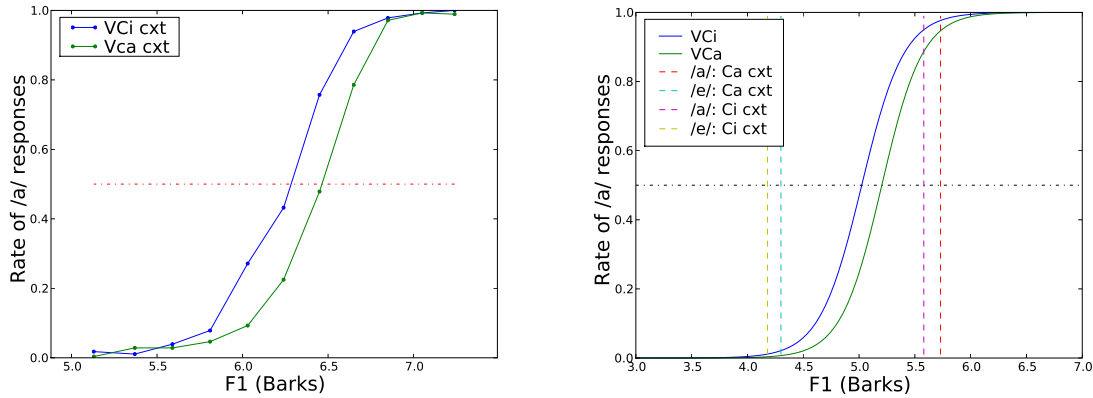


Figure 1: Left: Observed V_1 identification function in $V_2=/a/$ and $V_2=/i/$ contexts. Right: Predicted identification functions. Vertical lines are observed category means, horizontal lines indicate the crossover point (rate=0.5).

Observed

- Slightly-modified replication of Beddor et al. (2002) English study.
- 2AFC on V_1 in V_1CV_2 context, $C=/p/$ or $/b/$, $V_2=/a/$ or $/i/$.
- Identification curves shown in L of Fig. 1.
- From similar results, Beddor et al. argue native English listeners perceptually compensate for anticipatory vowel-to-vowel coarticulation in VCV sequences.

Expected

- Identification curves from (2) plotted in R of Fig. 1 for $/a/$ and $/i/$ contexts.
- μ , σ parameters taken from productions of adV_1CV_2 sequences ($V_1 \& V_2=/a/$, $/e/$, or $/i/$ and $C=/p/$ or $/b/$) by 6 native English speakers.²
- Means, standard deviation of F1 for V_1 in V_2 context calculated for four cases: $V_1=/a/$ or $/e/$, $V_2=/a/$ or $/i/$. Category frequencies for all cases were estimated from CELEX.
- Parameters, where c_1 is “ $V_1=/a/$ ”, c_2 is “ $V_1=/e/$ ”. All means and variances in Barks:³

V_2	μ_{c_1}	μ_{c_2}	$\sigma_C^2 + \sigma_S^2$	f_2/f_1
$/a/$	5.73	4.30	0.261	2.82
$/i/$	5.58	4.18	0.264	2.14

5. Predictions

- RA account of PC allows us to reason, make predictions about sound change/phonologization.
- Example:
 - Ohala (1993), etc.: Sound changes based in listener misperceptions, e.g. failure to compensate \rightarrow phonologization of a phonetic pattern.
 - However, sound change without loss of conditioning environment common.

²Repeated 10 times in random order. Formant values were measured at the midpoint of the target V_1 .

³Variance for $V_1=/a/$ taken to be the mean of the variance for aCa stimuli and the variance for aCi stimuli, similarly for $V_1=/e/$.

- Proposed RA account:
 - * Observed PC effect corresponds to different values of S' (crossover point) in contexts k_1 and k_2 .
 - * The second term of (3) predicts that $S'(k_1)$ and $S'(k_2)$ depend on the relative frequencies of c_1 and c_2 in these contexts.
 - * \implies if f_1/f_2 differs significantly by context, the PC effect can be diminished or canceled out.
 - * \implies failure to compensate could occur for sudden change in f_1/f_2 for k_1 but not k_2 , e.g. from an unrelated merger, word frequency drift..
- From (3), Compensation could also be undone by change in variances ($\sigma_C^2 + \sigma_S^2$) or category mean differences ($\mu_{c_1,k} - \mu_{c_2,k}$) for k_1 vs. k_2

6. Discussion

- RA is a different type of explanation from existing theories of PC: the analysis is of the cognitive task rather than the mechanism by which the task is achieved.⁴
- RA provides the most general expression of the task involved in PC effects. While RA might provide useful guidance in the proposal of possible mechanisms, it is itself not bounded by such considerations (Anderson, 1990).
- Model proposed here:
 1. Is compatible with exemplar-based theories, which have so far not dealt with PC effects.
 2. Allows incorporation of both speech-specific and general auditory factors.
 3. Very simple and only a first step toward understanding perception in general.
- Future work:
 - Develop more sophisticated rational models, e.g. unequal category variances, multiple categories.
 - Explore effects on predicted categorization behavior.

References

- Anderson, J. (1990). *The Adaptive Character of Thought*. Erlbaum, Hillsdale, N.J.
- Beddor, P., Harnsberger, J., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591–627.
- Beddor, P. and Krakow, R. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106:2868–2887.
- Elman, J. and McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27(2):143–165.
- Feldman, N. and Griffiths, T. (2007). A rational account of the perceptual magnet effect. In McNamara, D. and Trafton, J., editors, *Proceedings of the 29th Annual Cognitive Science Society*, pages 257–262, Austin, TX. Cognitive Science Society.
- Fowler, C. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99:1730–1741.
- Fowler, C. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68(2):161–177.
- Fowler, C., Best, C., and McRoberts, G. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48(6):559–70.
- Fried, L. and Holyoak, K. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(2):234–257.

⁴Alternatively, it is closer to the *computational* than the *algorithmic* or *implementational* levels of Marr (1982).

- Kingston, J. (1992). The phonetic and phonology of perceptually motivated articulatory covariation. *Language and Speech*, 35:99–114.
- Kingston, J. and Diehl, R. (1995). Intermediate properties in the perception of distinctive feature values. In Connell, B. and Arvaniti, A., editors, *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, pages 7–27. Cambridge University Press, Cambridge, UK.
- Lotto, A. (2004). Perceptual compensation for coarticulation as a general auditory process. In Agwuele, A., Warren, W., and Park, S.-H., editors, *Proceedings of the 2003 Texas Linguistics Society Conference*, pages 42–53, Somerville, Mass. Cascadilla.
- Lotto, A. and Kluender, K. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, 60(4):602–19.
- Lotto, A., Kluender, K., and Holt, L. (1997). Perceptual compensation for coarticulation by Japanese Quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102:1134–1140.
- Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: evidence from Japanese listeners’ perception of English “l” and “r”. *Cognition*, 24(3):169–96.
- Marr, D. (1982). *Vision*. W.H. Freeman, San Francisco.
- Ohala, J. (1993). The phonetics of sound change. In Jones, C., editor, *Historical Linguistics: Problems and Perspectives*, pages 237–278. Longman, London.
- Samuel, A. and Pitt, M. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48(2):416–434.
- Tenenbaum, J. and Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24:629–640.

Appendix A: Derivation of Eqn. 2

To evaluate (1), first integrate over T to find the $P(S | c_i, k)$:

$$P(S | c_2, k) = \int P(S | T, c_2, k) P(T | c_2, k) dT \quad (4)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_c} \frac{1}{\sqrt{2\pi}\sigma_S} \int \exp\left[-\frac{1}{2} \left(\frac{(\mu_{c_2,k} - T)^2}{\sigma_C^2} + \frac{(S - T)^2}{\sigma_S^2} \right)\right] dT \quad (5)$$

The argument of the integral is

$$\frac{(\mu_{c_2,k} - T)^2}{\sigma_C^2} + \frac{(S - T)^2}{\sigma_S^2} = \frac{T^2 - 2\mu_{c_2,k}T + \mu_{c_2,k}^2}{\sigma_C^2} + \frac{T^2 - 2ST + S^2}{\sigma_S^2}.$$

Complete the square in T :

$$\begin{aligned} \frac{T^2 - 2\mu_{c_2,k}T + \mu_{c_2,k}^2}{\sigma_C^2} + \frac{T^2 - 2ST + S^2}{\sigma_S^2} &= T^2 \left(\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2} \right) - 2T \left(\frac{\mu_{c_2,k}}{\sigma_C^2} + \frac{S}{\sigma_S^2} \right) + \left(\frac{\mu_{c_2,k}^2}{\sigma_C^2} + \frac{S^2}{\sigma_S^2} \right) \\ &= \left(\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2} \right) \left(T^2 - 2T \left(\frac{\frac{\mu_{c_2,k}}{\sigma_C^2} + \frac{S}{\sigma_S^2}}{\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}} \right) + \left(\frac{\frac{\mu_{c_2,k}^2}{\sigma_C^2} + \frac{S^2}{\sigma_S^2}}{\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}} \right) \right) \\ &= \left(\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2} \right) \left(T - \frac{\frac{\mu_{c_2,k}}{\sigma_C^2} + \frac{S}{\sigma_S^2}}{\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}} \right)^2 + \left(\frac{\mu_{c_2,k}^2}{\sigma_C^2} + \frac{S^2}{\sigma_S^2} \right) - \left(\frac{\frac{\mu_{c_2,k}}{\sigma_C^2} + \frac{S}{\sigma_S^2}}{\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}} \right)^2 \\ &= \alpha(T - \beta)^2 + \gamma \end{aligned} \quad (6)$$

Where

$$\alpha = \frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}, \quad \beta = \frac{\frac{\mu_{c_2,k}}{\sigma_C^2} + \frac{S}{\sigma_S^2}}{\frac{1}{\sigma_C^2} + \frac{1}{\sigma_S^2}}, \quad \gamma = \frac{1}{\sigma_C^2 \sigma_S^2} \left(\mu_{c_2,k}^2 \sigma_S^2 + S^2 \sigma_C^2 - \frac{(\mu_{c_2,k} \sigma_S^2 + \sigma_C^2 S)^2}{\sigma_S^2 + \sigma_C^2} \right)$$

Now, since

$$\int_{-\infty}^{\infty} e^{-A(x-B)^2+C} dx = e^C \sqrt{\frac{\pi}{A}} \quad (\text{for } A > 0)$$

substituting (6) into (5) and integrating gives

$$P(S | c_2, k) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_c^2 + \sigma_S^2}} \exp\left[-\frac{1}{2\sigma_C^2\sigma_S^2}(\mu_{c_2,k}^2\sigma_S^2 + S^2\sigma_C^2 - \frac{(\mu_{c_2,k}\sigma_S^2 + \sigma_C^2 S)^2}{\sigma_S^2 + \sigma_C^2})\right] \quad (7)$$

The ratio of (7) for c_2 and c_1 is then

$$\frac{P(S | c_2, k)}{P(S | c_1, k)} = \exp(b - Sg) \quad (8)$$

after some algebra, with

$$b = \frac{1}{2} \frac{\mu_{c_1,k}^2 - \mu_{c_2,k}^2}{\sigma_S^2 + \sigma_c^2}, \quad g = \frac{\mu_{c_1,k} - \mu_{c_2,k}}{\sigma_c^2 + \sigma_S^2}$$

Eqn. 1 can be rewritten as

$$P(c_1 | S, k) = \left(1 + \frac{P(c_2 | k) P(S | c_2, k)}{P(c_1 | k) P(S | c_1, k)}\right)^{-1} \quad (9)$$

Substitute (8) back into (9), defining $f_1 = P(c_1 | k)$, $f_2 = P(c_2 | k)$:

$$P(c_1 | S, k) = \left(1 + \frac{f_2}{f_1} e^{b-Sg}\right)^{-1}$$

This is Eqn. 2.