Using probability distributions to account for recognition of canonical and reduced word forms

Meghan Clayards, Centre for Research on Language Mind and Brain, McGill University, Montreal, Quebec Canada, meghan.clayards@mcgill.ca

ABSTRACT: The frequency of a word form influences how efficiently it is processed, but canonical forms often show an advantage over reduced forms even when the reduced form is more frequent. This paper addresses this paradox by considering a model in which representations of lexical items consist of a distribution over forms. Optimal inference given these distributions accounts for item based differences in recognition of phonological variants and canonical form advantage.

Reduction: Variation in speech is ubiquitous. This paper deals with reduction – variation in the pronunciation of word forms due to a casual speaking style or a fast speaking rate. Reduction is highly prevalent in conversational speech (Ernestus, Baayen, & Schreuder, 2002). Johnson (2004) found that as much as 22% of multi-syllable words had at least one syllable deleted; a very extreme but common form of reduction. A number of context-dependant and context-independent factors influence whether and how much reduction occurs including predictability or frequency of the word, speaker gender, speaking style and rate (Aylet & Turk, 2006; Bell et al., 2009; Bybee 2001; Pluymaekers, Ernestus & Baayen, 2005; Raymond, Datricourt & Hume, 2006). Thus, words differ in *how likely* they are to reduce – for example Ranbom and Connine (2007) found that 'centre' is overwhelmingly produced with a nasal flap as in 'cenner', while 'dental' is less often produced with a flap and 'mental' even less frequently contains a flap. Words also differ in *how much* they reduce – for example Lavoie (2003) found that the word 'for' reduced much more than its homophone 'four' and had many more different reduced forms. Thus words don't just have different allomorphs of reduced and unreduced forms, they have a range of acoustic-phonetic forms and the range is specific to the lexical item, not just to the phonological form.

Word Recognition: Research on the recognition of variant forms has revealed two apparently contradictory phenomena. The first is sensitivity to variant frequency – listeners are very sensitive to how often a particular variant of a word is produced and process more frequent forms more efficiently than less frequent forms (Dautricourt, 2004; Ranbom, Connine & Yudman, 2009) even when word frequency is taken into account (Ranbom & Connine, 2007). The second is the apparent special status of the canonical form – carefully articulated or canonical forms are often processed faster than even more common variants (Ranbom & Connine, 2007) even when presented in the original phonetic context (Dautricourt, 2004). Canonical forms have been shown to produce more phoneme restoration and a larger lexical bias (Pitt, 2009) as well as more long-term priming (Sumner & Samuel, 2005) even though they are the less frequent form.

Models of recognition: Models of word recognition do not typically deal well with variation. Models such as TRACE (McClelland & Elman, 1986) or Shortlist (Norris, 1994) assume canonical forms and leave the variation to be dealt with at the sub-lexical level and models of sub-lexical fix up (Gaskell & Marslen-Wilson 1996; Gow, 2003; Mitterer, Csepe, Honbolygo, & Blomert, 2006) which uncover canonical forms from the changed or reduced signal, can not account for variant frequency effects. On the other hand, in models that are built to account for variant effects – e.g exemplar models (Goldinger, 1998; Hawkins, 2003; Pierrehumbert, 2003) which have a cloud of exemplars rather than a single representation – variant frequency is explicitly represented. It is less obvious, however, how one could account for the advantage of an *infrequent* canonical form. A third solution is to include both a canonical form and graded representations of reduced forms in the lexical representation (Ranbom & Connine, 2007). However, this fails to capture the whole range of reduced forms and multiple representations have been somewhat unsuccessful when implemented (Scharenbourg & Boves, 2002).

What is needed is a model that captures lexically specific ranges of variability and can account for both variant frequency effects and the advantage of the canonical form. As a solution this paper proposes that lexical representations should be thought of as probability distributions over a range of continuous phonetic forms as in Pierrehumbert (2003). To examine word recognition a simple Bayesian model is used in which the posterior probability of potential lexical candidates is evaluated. This approach allows one to examine the properties of a distributional model without having to model the entire lexicon explicitly. However, many other models including exemplar models would have the same properties. The following sections will argue that this model has the desired properties, namely sensitivity to variant frequency and the advantage for canonical forms.

Probability distributions: We begin by conceptualizing lexical representations of form as probability distributions over a range of phonetic forms. For simplicity Gaussian probability density functions (pdfs) varying along a single phonetic dimension are used. Figure 1a shows hypothetical pdfs for a set of words varying in form from nasal-oral stop clusters to nasal flaps to true nasals.

Each word has its own distribution reflecting the range and frequency of productions observed. For example, 'rental' which reduces often would have a distribution at the nasal end of this continuum while 'gentle' which is less often reduced would have a distribution at the nasal-oral stop end of the continuum. From this figure it is easy to see the difference in probability between reduced forms of different words. To understand why the infrequent canonical forms would have any special status given these representations it is important to note that canonical forms represent the extreme end of each distribution.

A parallel phenomenon exists in the perception of phonetic categories. Listeners rate extreme exemplars more highly than less extreme but more frequent productions (Johnson, Flemming & Wright, 2003).

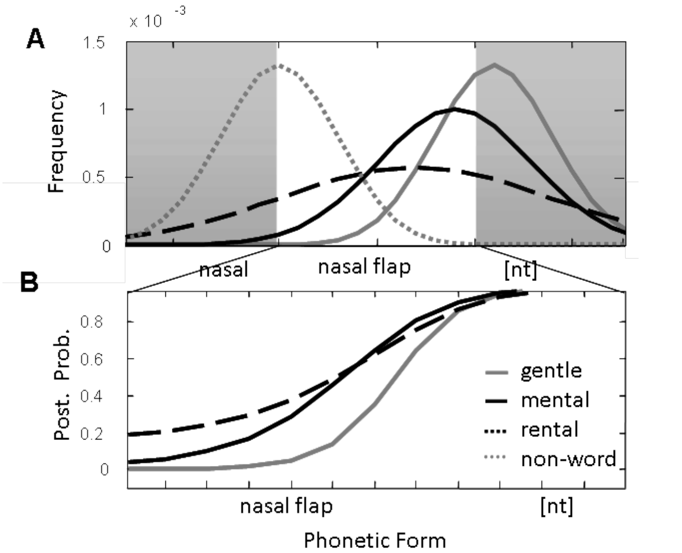

Figure 1: [A] Hypothetical probability distributions for three words varying in the degree of reduction, and one distribution representing non-word competitors with a true nasal. [B] Probability of each word given each phonetic form calculated from the equation in (1). The region of interest on the continuum of A has been expanded for B.

One explanation for this phenomenon is that because there is always some overlap between the phonetic characteristics of categories, the most common productions may not be the most distinctive or least confusable (Boersma, 2006). In Boersma's Stochastic Optimality account, the most distinctive production wins over the most frequent because it is the least confusable. The same prediction is made by a Bayesian model of word recognition outlined below.

Posterior probability: We can use the equation in (1) from Bayes rule and our probability distributions to calculate the posterior probability of each word given a phonetic form and the potential competitors (p(word|form). For simplification purposes we will only consider one competitor, a non-word which is identical to the word in all aspects except that it contains a true nasal (as pictured in Figure 1a). This allows us to consider just variation along this single dimension. All other words and non-words would differ along other phonetic dimensions as well and those dimensions would have to be taken into consideration to make the proper comparison. It is also appropriate for modeling a lexical decision task in which the listener decides whether they have heard a word or a non-word.

(1) $$p(word \mid form) = \frac{p(form \mid word)\,p(word)}{p(form \mid word)\,p(word) + p(form \mid nonword)\,p(nonword)}$$

The probability of the phonetic form given the word (p(form|word)) is represented by the height of the corresponding line in Figure 1 at the point on the x axis corresponding to the form. This means that phonetic forms that occur often for a word will have higher posterior probability than forms that occur less often. The probability of the word (p(word)) will depend on its predictability given the context and its frequency in general with higher posterior probability for more frequent or predictable words. In this model frequency and predictability do not affect reduction directly, but more frequent or predictable forms will have higher posterior probability and will therefore be more easily recognized with ambiguous or reduced pronunciations. The results reported here make the simplifying assumption that all words and non-words are equally frequent in a lexical decision task. The denominator of the equation includes the same computation for all potential competitors. Thus forms that may not occur very frequently with a word can still have high posterior probability if they occur even less frequently with any other word – i.e. they are *distinctive*.

Modeling RT data: Posterior probability itself is not a behavioral measure. However there is reason to believe that it might relate directly to behavioral measures in word recognition tasks. Norris (2006; 2009) provides a complete model of how reaction time and posterior probability might be related in lexical decision and word recognition. In general higher posterior probability means faster recognition. Further evidence comes from a study which explicitly manipulated the probability of different phonetic forms in a word recognition task and found that looking times and response choices closely followed the posterior probability given these forms (Clayards, Tanenhaus, Aslin & Jacobs, 2008). Further analysis found a similar pattern for reaction time (Clayards, 2008).

Figure 1b shows the posterior probability of each of the words from Figure 1a for each of the phonetic forms, calculated using the equation in (1). Words that reduce more often (rental) have a higher posterior probability for the reduced forms than words that reduce less often (gentle). Importantly, forms towards the canonical end of the continuum also have high posterior probability. This is because although they do not occur often with e.g. "rental", the would occur even less often with the potential non-word competitor "rennal". These simulations predict that reduced forms should be processed more efficiently the more often they occur, and that forms towards the canonical end should also be processed efficiently.

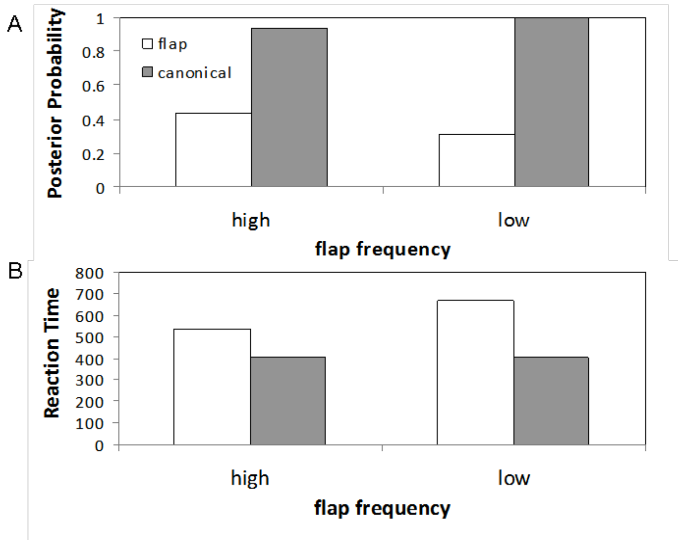

Figure 2: [A] Posterior probability for each set of words calculated from the model and corpus frequencies. [B] Lexical decision reaction times from Ranbom & Connine 2008 for the same sets of words.

In order to provide a more quantitative model, corupus data on transcribed relative frequencies of flapped and canonical forms and lexical decision reaction times to those forms from Ranbom and Connine (2007) (R&C) were used. Likelihoods (p(form|word)) for each form and each word were calculated from the relative frequencies reported in the corpus experiment (instances of form/total number of tokens from Appendix A, R&C 2007). The likelihood of each form given the non word was assumed to be very low for the canonical form (0.001/150) and high for the reduced form (149.999/150). The lexical decision data in R&C is divided into low flap frequency words and high flap frequency words. Average posterior probability was calculated for these two groups of words for the flapped and the canonical forms. Figure 2 shows the average RTs from R&C and the average posterior probabilities from the model. The model predicts lower reaction time (RT) for higher posterior probability, so lower RT for canonical than flapped forms and lower RT for high versus low flap frequency for the reduced forms. This is the pattern observed in the RT data.

Summary: This paper presents a simplified model of probabilistic word recognition in which lexical representations of form are probability distributions over a range of phonetic forms and the probability of each candidate is computed given a particular form. Only one phonetic dimension and only words and non-words that differed in that dimension were considered. A full model would have to take into account multiple dimensions and real competitor sets. None the less, this model demonstrates that, in principle, such a model can account for both frequency effects of reduced or variant forms and the advantage of canonical forms. This is because, once all dimensions are considered, canonical forms will be tend to be less like competitors than reduced forms.

The success of this approach will depend on the extent to which this simplified model is representative of the larger model. That is, to what extent are canonical forms actually more distinct from other lexical items.

References:

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *JASA, 119*(5), 3048-3054.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *JML, 60*(1), 92-111.

Boersma, P. (2006). Prototypicality judgments as inverted perception. In G. Fanselow, C. Fery, R. Vogel & M. Schlesewky (Eds.), *Gradience in grammar: Generative perspectives* (pp. 167-184). Oxford: Oxford Linguistics.

Bybee, J. (2001) Phonology and Language Use. Cambridge University Press, Cambridge UK.

Clayards, M. (2008). The ideal listener: Making optimal use of acoustic-phonetic cues for word recognition. Unpublished Doctoral Thesis.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Speech perception reflects optimal use of probabilistic speech cues. Cognition, 108, 804-809.

Dautricourt, R. (2004). *Word boundary palatalization in English and its effects on word recognition*. Poster presented at the meeting of the LSA, Boston, MA.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1-3), 162-173.

Gaskell, G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *JEP:HPP, 22*, 144-158.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psych Review, 105*(2), 251-278.

Gow, D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *P&P, 65*(4), 575-590.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phonetics, 31*(373-405).

Johnson, K. (2004). *Massive reduction in conversational speech*. Tokyo: The National Institute for Japanese Language.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*(3), 505-528.

Lavoie, L. (2003). Some influences on the realization of for and four in American English. *JIPA, 32*(02), 175-202.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cog Psych, 18*(1), 1-86.

Mitterer, H., Csepe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimliated words does not depend on specific language experience. *Cog Sci, 30*(3), 451-471.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327-357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. Psychological Review, 116(1), 207-219.

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), Probability theory in linguistics. Cambridge, MA: MIT Press.

Pitt, M. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *JML, 61*, 19-36.

Ranbom, L., & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *JML, 57*(2), 273-298.

Ranbom, L., Connine, C., & Yudman, E. (2009). Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *JEP: HPP, 35*(4), 1205-1220.

Raymond, W., Dautricourt, R., & Hume, E. (2006). Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Lang Var & Change, 18*(01), 55-97.

Scharenborg, O., & Boves, L. (2002). *Pronunciation variation modelling in a model of human word recognition.* In ISCA Tutoral and Research Workshop. Pronunication Modeling and Lexical Adaptation for Spoken Language. Estes Park, Colorado

Sumner, M., & Samuel, A. (2005). Perception and representation of regular variation: The case of final/t. *JML, 52*(3), 322-338.

Pluymaekers, M., Ernestus, M., & Baayen, R. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *JASA, 118,* 256.