NATURALISTIC SPEECH MISPERCEPTION – A COMPUTATIONAL CORPUS-BASED STUDY

Introduction: Laboratory studies, historical linguistics and theoretical modelling have elucidated a number of phonetic trends of misperception. However, little work to date has investigated misperception in its most naturalistic form, namely slips of the ear (Bond, 1999), e.g. "End Rule Left" → "Andrew left", "Geez, really?" → "Disraeli". Our goal in this research is to develop a computational analysis followed by quantitative statistics based on a 3500+ pair corpus of slips of the ear. Investigation the targets and directionality of misperception in the "messy" realm of real conversation can provide a testbed for numerous theoretical and experimental constructs, as applied to speech perception (Miller and Nicely, 1955), word segmentation from phrases (Mattys et al., 2005) and cross-dialectal comprehension (Labov, 2010).

Methods: As the old adage goes, "Data is not the plural of anecdote", and thus to draw reasonable and reliable conclusions about whether naturalistic occurrences of this sort end up mirroring, paralleling, or diverging from independent evidence collected with laboratory methods, we have collected and compiled a corpus of 3,638 naturalistically occurring instances (the largest existing corpus to our knowledge), consisting of slips from English spontaneous conversation, mostly North American varieties phonetically transcribed using collectors' transcriptions and interlocutors' demographics.

Analysis of the errors in terms of consonant substitutions, deletions, or insertions was computationally extracted using alignment algorithms from computational biology (Kondrak, 2003), with subsequent application of parameter optimization techniques, the ultimate outcome of which was a two-dimensional confusion matrix of all substitutions (treating insertion or deletion as alignment with zero).

Results: A frequency analysis of deletions and substitutions of consonants yield the following trends (normalised rate using corpus token frequencies), in terms of a scale of confusability.

["≻" indicates "is confused with a higher rate than"]

(1)	Adjacent environments (Deletions)	C_C ≻ C_V ≻ V_C ≻ V_# ≻ #_V ≻ V_V (see Figure 1)
(2)	Adjacent environments (Substitutions)	V_V ≻ #_V ≻ V_# ≻ C_V ≻ V_C ≻ C_C (see Figure 1)
(3)	Place (Deletions)	Coronal ≻ Dorsal ≻ Labial
(4)	Place (Substitutions)	Labial/Dorsal ≻ Coronal

The normalised trends in adjacent environments for deletions (1) and substitutions (2) came out to be significant against a hypothetical trend based on increasing phonetic cues and inspecting the role of adjacency alone, C_C ≻ C_V ≻ V_C ≻ V_# ≻ #_V ≻ V_V, in a logistic regression - using *lmer()* in R with helmert coding in either directions, which compares each level to the mean of the subsequent levels (forward) or previous levels (reversed), at $.001 < p < .05$ for most of the levels of contrasts.

Substitution/deletion complementarity of environments: For deletions, the environment trend (1) provides a negative correlation between the amount of confusion with the amount of phonetic cues, such that phonetically less robust environments would have more confusions, where C_C has the highest confusion rate. However this correlation is reversed for substitutions (2), where V_V has the highest confusion rate. Together, (1) and (2) indicate a substitution/deletion complementarity (see Figure 1), which could be explained by considering that in more robust environment (e.g. V_V), the fact that a segment was there must be retained, and so errors, when they occur, are more likely ones of substitution.

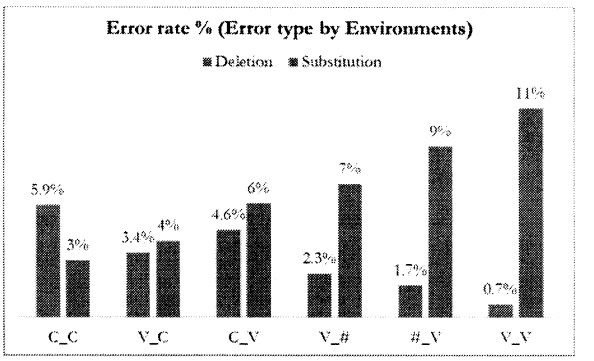

Figure 1

Scale of least perceptibility of Place: The place trend for deletions (3) matches largely with Jun's (2004) scale of least perceptibility, Coronal ≻ Labial ≻ Velar. In a context-sensitive analysis, the divergence of Dorsal being more confusable than Labial turned out to be caused mainly by /k/ deletions in V_C[sibilant], e.g. "section" → "session".

Underspecification of Coronal: The place trend for substitutions (4) shows that coronal confusion is significantly lower than Labial ($p < .001$) and Dorsal ($p < .01$). In fact, a more specific comparison, by excluding confusions between Labial and Dorsal, shows that the probability of (Dorsal/Labial → Coronal) is significantly higher than the probability of (Coronal → Dorsal/Labial), at $p < .01$. These analyses can be captured with a featurally underspecified lexicon (FUL) (Eulitz & Lahiri, 2004) such that Coronal is underspecified or with an acoustic/perceptual account by asymmetric distances between coronal and non-coronal (Tsuji, Cristia, & Fikkert, 2012).

Stressed syllables - "islands" of reliability: Stressed syllables are considered as "islands" of reliability (Pisoni's term) with greater prominence, duration and intensity, thus a plausible hypothesis would be unstressed syllables being more confusable than stressed. However, as opposed to our hypothesis, an initial analysis of the effect of stress with pre/post-vocalic environments, #_V and V_#, shows that stressed syllables are more confusable, but only in monosyllables and not in polysyllables. (Browman, 1980)

Phonological/phonetic processes: Finally, certain substitution-pairs showed asymmetric confusions which suggest that phonotactic processes specific to the phonology of English phonological/phonetic processes play a role, namely, θ-fronting, ŋ-alveolarization and t-d-neutralisation due to tapping ([d] ↔ [ɾ]). Surprisingly, most of these asymmetric confusions are also found in laboratory studies with nonsense stimuli for example in Miller and Nicely (1955), Wang and Bilger (1973) and Cutler et al. (2004).

Conclusions: We reported new findings using the largest naturalistic speech misperception corpus and alignment algorithms. We explored the effects of positional environments, stress information, and phonological features and how perceptual errors might drive diachronic changes. We showed that naturalistic data are consistent and complementary with laboratory results, although much richer because they include many more phonological contexts than laboratory studies (usually limited to VCV or CV). We demonstrated that phonological and perceptual considerations exert a major role in real-life messy, everyday erroneous performance even when we might expect top-down contextual effects to otherwise dominate. Given the divergences from our hypotheses and the different patterns observed for deletions and substitutions, we outline how context-sensitive confusion matrices could refine these conclusions.

References:

Bond, Z. 1999. *Slips of the ear: Errors in the perception of casual conversation.* Academic Press.

Browman, C.P. 1980. Perceptual processing: Evidence from slips of the ear. *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* 213–230.

Cutler, A., A. Weber, R. Smits, and N. Cooper. 2004. Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America* 116:3668.

Eulitz, C. & Lahiri A. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16:4, 577-583.

Jun, J. 2004. Place assimilation. In *Phonetically based phonology* , ed. B. Hayes, R.M. Kirchner, and D. Steriade, 58–86. Cambridge Univ Pr.

Kondrak, G. 2003. Phonetic alignment and similarity. *Computers and the Humanities*, 37(3), 273--291.

Labov, William. 2010. *Natural misunderstandings*, chapter 2, 19–47. Wiley-Blackwell. URL http://dx.doi.org/10.1002/9781444327496.ch2.

Mattys, S.L., L. White, and J.F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology*-General 134:477–500.

Miller, G.A., and P.E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America* 27:338.

Tsuji, S., Cristia, A., & Fikkert, P. 2012. '/ompa/ is further away from /onta/ than vice versa for Dutch infants: Prelexical bases of the labial-coronal perceptual asymmetry'. Oral presentation at the Manchester Phonology Meeting 2012, Manchester, UK.

Wang, M.D., and R.C. Bilger. 1973. Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America.*