Effects of context-sensitive phonetic variation and lexical structure on the uniqueness of words.

Edward T. Auer, Jr. and Lynne E. Bernstein

Spoken Language Processes Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057

Abstract: Phonetic context can affect speechreading confusions for phonemes. In Experiment I, behavioral experiments were performed to examine effects of context-sensitive phonetic variation on the visual confusability of consonants and vowels. In Experiment II, computational experiments were performed to assess the importance of patterns of context-sensitive visual confusability on the uniqueness of words in the language. Results from Experiment I further support the conclusion that phonetic context influences phoneme confusability. The computational experiments in Experiment II provide evidence that the distribution of words in English substantially preserves lexical uniqueness even when phonetic variability is taken into account.

INTRODUCTION

Previously, (1) investigated the relationship between visually speechreadable phonemic distinctions and the predicted uniqueness of speechread words in English. (1) demonstrated that the distribution of words in English substantially preserves lexical uniqueness, and that estimates of lexical uniqueness are sensitive to small changes in the number of available phonemic distinctions. For example, the loss of the phonemic distinctions among /b/, /p/, and /m/ results in a loss of lexical uniqueness for the words "bat," "pat," and "mat." However, the word "bought" remains unique, because "pought" and "mought" are not words in English. In (1), estimates of phonetic similarity were based on phoneme identifications in a single phonetic context(Consonant /a/ and /h/ Vowel /g/). The current study extends this work by including effects of coarticulation in estimates of phonetic similarity.

EXPERIMENT I

Coarticulation effects arising from variation in surrounding phonetic contexts have been demonstrated to alter phoneme identification by speechreaders (2,3,4). In this experiment, effects of context-sensitive phonetic variation on the identification of consonants and vowels were examined. Subjects were 10 severe-to-profound, congenitally hearing-impaired adults, aged 18-30 years, with English as a first language, 20/30 or better vision in both eyes, and average or better speechreading ability. Five subjects participated in each condition.

The stimulus set included **initial consonants and clusters** (two tokens each of /b p m f v ð θ tʃ dʒ ʃ w r d h g k l n s z t j pr st tr gr kr/ spoken in the four contexts, C-/adəd/, C-/idəd/, C-/udəd/, and C-/ədad/) and **vowels** (two tokens each of the vowels /i ɪ ɛ æ a ɔ ʌ ʊ u/, r-colored vowels / ɝ ɪr ʊr ar er/, and diphthongs /eɪ oʊ aʊ aɪ ɔɪ/, spoken in the four contexts, /m/-V-/m/, /n/-V-/n/, /p/-V-/p/, and /t/-V-/t/). A female adult, native speaker of English was professionally videotaped in color against a neutral background with her head filling the screen. The stimuli were stored on optical videodisks and were presented on a 14-inch color monitor.

Each stimulus set was presented in randomized order a total of 10 times. Subjects were instructed to identify the target phoneme in the spoken nonsense word and to indicate their choice by pressing the appropriately labeled key on the keyboard in front of them. Feedback was provided on all trials.

TABLE 1. Percent correct for consonants and vowels as function of context.

	C-/adəd/	C-/idəd/	C-/ədad/	C-/udəd/	/m/-V-/m/	/p/-V-/p/	/n/-V-/n/	/t/-V-/t/
%Correct	43	36	31	35	75	76	75	74

Vowel identification was more accurate than consonant identification. Furthermore, vowel identification accuracy did not vary as a function of the surrounding consonantal context. However, consonant identification accuracy varied as a function of vowel context. Consistent with previous studies, consonant identification accuracy was low in the /u/ environment, a likely result of lip rounding for /u/. These results suggest that accurate estimates of phonetic confusability should take into account phonetic environment of the phonemes.

EXPERIMENT II

In Experiment II, three models of optical phonetic similarity were used to computationally assess the importance of patterns of context-sensitive visual confusability on the uniqueness of words in the language. In Model 1, consonants in all positions of words were transcribed as if they were as intelligible as consonants before the vowel /a/. In Model 2, consonants in all positions were transcribed as if they were as intelligible as consonants before the vowel /u/. In Model 3, consonants in syllables containing the vowels /u,ɝ,ʊ,ʊr/ were transcribed using model 2, and all other consonants were transcribed using Model 2.

Computational lexical modeling techniques (1,5,6) were applied as follows: First, a phonemically transcribed computer-readable lexical database, PhLex, (7) was selected to serve as a representative sample of words in English. Second, transcription rules were defined in the form of symbol substitutions for all phonemes in phonemic equivalence classes. A phonemic equivalence class comprised the set of phonemes or clusters modeled as mutually confused using the behavioral data from Experiment I. Phonemic equivalence classes for vowels and for consonants as a function of vowel context were as follows: **VOWELS:** {i,ɪ}, {ɛ,æ}, {a, ʌ, ə}, {ɔ,ar,ɔr}, {ɝ,ʊ,ʊr}, {u}, {ɪr}, {er}, {eɪ}, {oʊ}, {aʊ}, {aɪ}, {ɔɪ}; **INITIAL /a/:** {b,p,m}, {pr}, {f,v}, {ð,θ}, {ʃ,tʃ,ʒ,dʒ}, {w}, {d,t,s,z,n,k,ŋ,g,j,st}, {h}, {r,gr,kr}, {l}, {tr}; **INITIAL /u/:** {b,p,m,pr},{f,v}, {ð,θ}, {w,r,gr}, {l}, {ʃ,tʃ,ʒ,dʒ,d,t,s,z,n,k,h,ŋ,g,j,st,tr,kr}. Third, the lexical database was then transcribed by applying the transcription rules. Lexical equivalence classes were formed by collapsing across identically transcribed words. Finally, metrics were computed to compare the distribution of patterns in the newly transcribed lexicon against the distribution of patterns in the original lexicon. Frequency-weighted percent words unique estimated the extent to which unique words are encountered in everyday language. Frequency-weighted expected class size estimated the average size of the lexical equivalence classes encountered in everyday language. See (1) for a description of the calculation and rationale for these two metrics.

Appropriately modeled contextual variations in consonantal intelligibility do not reduce lexical uniqueness relative to a high visibility model (Model 3 versus Model 1). Lexical uniqueness is reduced when reductions in segmental intelligibility are over-applied (Model 2 versus Model 1) across the entire lexicon. Thus, the choice of a phonetic similarity model does matter for modeling the uniqueness of words.

TABLE 2. Percent unique words and expected class size as a function of transcription rule set.

Transcription Rule Sets	Percent Unique Words	Expected Class Size
Model 1: Initial /a/	57.7	4.25
Model 2: Initial /u/	49.6	8.16
Model 3: Initial/a/ and Initial /u/	57.1	4.37

ACKNOWLEDGMENT

This research was supported by a grant from the US National Institutes of Health (DC02107).

REFERENCES

1. Auer, E. T. Jr., and Bernstein, L. E., *J. Acoust. Soc. Am.* 102(6), pp. 3704-3710 (1997).
2. Bengeurel, A. P. and Pichora-Fuller, M. K., *J. Speech Hear. Res.* 25, pp. 600-607 (1982).
3. Jackson, P. L. in C. DeFillipo and D. Sims, Eds, *New Reflections on Speechreading.*, A.G. Bell Association for the Deaf, Washington, DC, 1988 pp. 99-115.
4. Owens, E. and Blazek, B., *J. Speech Hear. Res.* 28, pp. 381-393 (1985).
5. Altmann, G. T. M. *Cognitive Models of Speech Processing*, Cambridge: MIT Press, 1990, ch. 10, pp. 211-235.
6. Carter, D., *Comput. Speech Lang.* 2, pp. 1-11 (1987).
7. Seitz, P. F., Bernstein, L. E., and Auer, E. T., Jr., *PhLex (Phonologically Transformable Lexicon), A 35,000-word computer readable pronouncing American English lexicon on structural principles, with accompanying phonological rules and word frequencies,* Gallaudet Research Institute, Washington, DC (1995).