Minneapolis, Minnesota
NOISE-CON 2005
2005 October 17-19

Frequency and Neighborhood Effects on Auditory Perception of Drug Names in Noise

Bruce L. Lambert
Swu-Jane Lin
Seng Wee Toh
Department Pharmacy Administration
833 S. Wood St., MC871, Rm. 271
University of Illinois at Chicago
Chicago, IL 60612-7231
lambertb@uic.edu

Paul A. Luce
Conor T. McLennan
Robert La Vigne
Department of Psychology
Park Hall, Room 206
University at Buffalo
Buffalo, NY 14260-4110

William M. Fisher
William M. Fisher Consulting
20265 Darlington Drive
Gaithersburg, MD 20886-1005

Laura Walsh Dickey
Department of Linguistics
2016 Sheridan Road
Northwestern University
Evanston, IL 60208-4090

John W. Senders
295 Indian Rd.
Toronto, Ontario M6R 2X5

ABSTRACT

Drug name confusions occur when drugs whose names look or sound alike (e.g., Celebrex/Cerebyx) are confused by clinicians and patients. Our project examines effects of frequency and similarity on auditory perception of drug names. Luce's Neighborhood Activation Model is the theoretical framework. Frequency information was taken from national prescribing frequency databases. Spoken drug names were transcribed to ARPAbet. Similarity neighborhoods were calculated based on phoneme edit distance using phoneme-to-phoneme confusion probabilities as substitution costs. The confusability of each name was estimated using Luce's frequency-weighted neighborhood probability rule (FWNPR). One hundred brand and one hundred generic drug names, stratified by frequency, were recorded. Low and high frequency sounds were filtered out to mimic telephone bandwidth. Names were used as stimuli in a perceptual identification experiment. Participants heard the drug names, at three different signal-to-noise ratios, against a background of 20-speaker babble. The task was to repeat back each name they heard. Participants were then shown the 190 names and asked to pronounce them and rate their familiarity. Data have been collected from 66 pharmacists. These data are being scored. Additional data will be collected from 50 physicians, 50 nurses, and 50 lay people. Results will be used to refine the theory and to build screening tools for drug companies and regulators.

1. INTRODUCTION

The Institute of Medicine (IOM) reports that between 44,000 and 98,000 Americans die each year as the result of medical errors, making such errors the fourth leading cause of death in the U.S. [1]. Errors involving medication are the most common type [1, 2]. One in six reported errors in the U.S. involves drug names that look or sound alike (e.g., *cisplatin/carboplatin, Retrovir/Ritonavir, Toradol/Tapazole*) [3-8]. In outpatient pharmacies, the wrong drug error rate is estimated to be 0.13% [9]. With more than 3 billion prescriptions dispensed annually in the U.S., this translates to more than 3.9 million wrong drug errors per year. The IOM recommended that: "The Food and Drug Administration (FDA) should…require pharmaceutical companies to

test (using FDA-approved methods) proposed drug names to identify and remedy potential sound-alike and look-alike confusion with existing drug names" [1]. The federal Quality Interagency Coordination Task Force (QuIC) committed within one year to "develop additional standards for proprietary drug names to avoid name confusion" (p. 26) [10]. Thus, the development of methods for minimizing drug name confusions addresses a significant public health problem. This article describes our work on the development of a theoretical framework and experimental protocol that will help us predict and prevent auditory drug name confusions. In section 2, we describe the theoretical model. In section 3, we describe how we operationalized the key quantities from the model, and in section 4 we describe our experimental methods. Since this is still work in progress, we do not yet have results to present. Instead, we conclude with a discussion of future plans and challenges.

2. BACKGROUND AND PROPOSED MODEL

To explain auditory perceptual confusions, we use Luce's Neighborhood Activation Model (NAM) [11]. NAM is an activation-competition model of auditory word perception [12, 13]. Figure 1 is a schematic description of NAM [11]. The metaphor underlying these models is that of spreading activation between networks of interconnected nodes. The nodes represent lexical (word) and sublexical (sub-word) units at various levels of abstraction. Nodes pass excitatory and inhibitory signals to one another over weighted connections. Each node has a level of activation that is a function of its resting activation as well as its excitatory and inhibitory inputs. Resting activation is itself a function of the frequency of occurrence of a given word or segment.

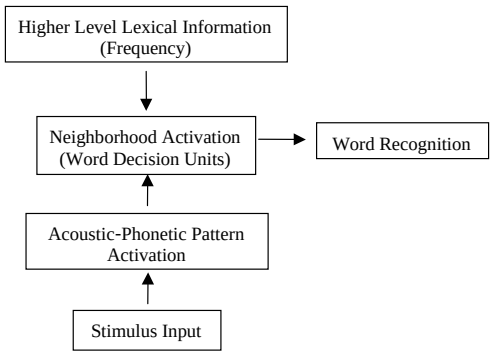

Figure 1: Flow chart for Luce's neighborhood activation model.

Word recognition begins when speech sounds activate sublexical, acoustic-phonetic patterns in working memory. These sublexical patterns correspond to phonemes or clusters of phonemes that match the input. Nodes corresponding to activated acoustic-phonetic patterns send excitatory signals to word decision units that contain the given patterns. Since most acoustic-phonetic patterns will be contained in many different words, a whole cohort of similar words will be activated, including the target word and all of its 'neighbors' [14, 15]. The *neighborhood* is a set of words that share one or more sublexical features. At the word level, each node sends inhibitory signals to each other word node. This pattern of mutual inhibition at the word level creates a competition among word nodes. Input to the word decision units also includes information about the frequency of occurrence of competing word candidates. The input to each word decision unit is combined to produce a word decision value. This value reflects the degree

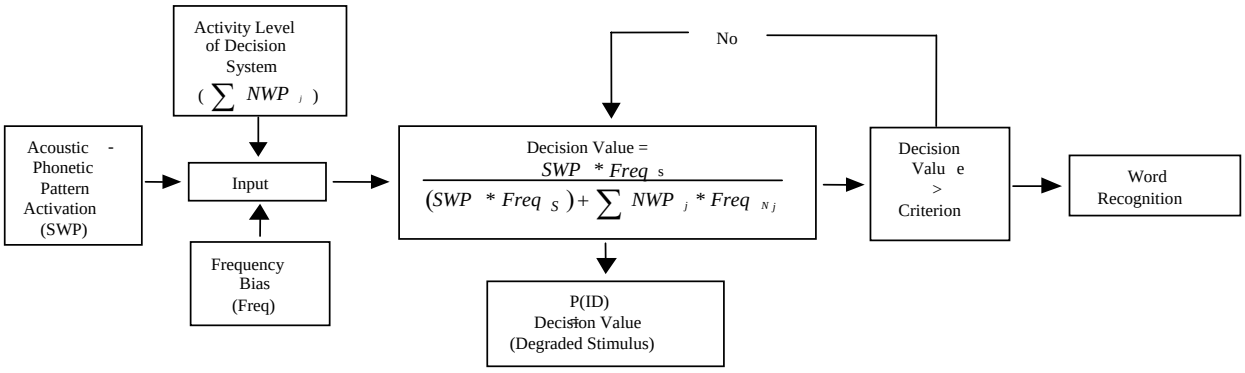

Figure 2: Diagram of a Single Decision Unit

of confidence in the hypothesis that a given word is in the input. When this value exceeds a threshold, recognition occurs. When stimulus information is incomplete or degraded by noise, it may be that no decision value exceeds the threshold. In this case, the decision value is interpreted as a probability. If a person is forced to identify a degraded word, the decision value, *p(ID)*, reflects the probability that a given word will be identified. The hypothetical structure of a word decision unit is shown in Figure 2 [11].

The NAM is a processing instantiation of Luce's frequency-weighted neighborhood probability rule (FWNPR), which is in turn based on R. D. Luce's forced choice rule [11, 16]. The forced-choice rule predicts that the probability of identifying a stimulus *S* is equal to the probability of *S* divided by the sum of the probability of *S* and the probability of *S*'s neighbors:

$$p(ID) = \frac{p(S)}{p(S) + \sum_{j=1}^{n} p(N_j)} \tag{1}$$

In NAM, *p(S)* is analogous to the conditional probability of the stimulus word given itself (i.e., the probability that a person will identify *Prozac*® when presented with *Prozac*®):

$$p(S) = SWP = \prod_{i=1}^{n} p(PS_i \mid PS_i) \tag{2}$$

where *SWP* stands for stimulus word probability and PS_i is the *i*th of *n* phonemes in the stimulus word. *SWP* can also be thought of as the segmental intelligibility of S. $P(N_j)$ from the forced choice rule becomes neighborhood word probability (*NWP*) in NAM:

$$p(N_j) = NWP = \prod_{i=1}^{n} p(PN_i \mid PS_i) \tag{3}$$

where $p(PN_i \mid PS_i)$ is analogous to the conditional probability of a neighbor word given the stimulus word (i.e., the probability that a person will identify *Prilosec*® when presented with *Prozac*®), and PN_i is the *i*th of *n* phonemes in the neighbor word. *NWP* can also be thought of as the segmental confusability of *S* and *N*. Each term in the rule is then weighted by frequency:

$$p(ID) = \frac{\left[\prod_{i=1}^{n} p(PS_i \mid PS_i)\right] * Freq_S}{\left[\prod_{i=1}^{n} p(PS_i \mid PS_i)\right] * Freq_S + \sum_{j=1}^{nn}\left\{\left[\prod_{i=1}^{n} p(PN_{ij} \mid PS_i)\right] * Freq_{N_j}\right\}} = \frac{SWP * Freq_S}{(SWP * Freq_S) + \sum(NWP * Freq_N)} \tag{4}$$

where *nn* is the number of neighbors and $Freq_{Nj}$ is the frequency of the *j*th neighbor. Note, however, that *SWP and NWP should not be interpreted literally as conditional probabilities*. *SWP* and the sum of all *NWP*s will not necessarily sum to 1, as they must if they were truly conditional probabilities. Rather, *SWP* and *NWP* are meant to model *activations* in the cognitive system [11, 17]. *P(ID)* is the only true probability, and it is a function of *SWP*, *NWP* and word frequencies.

Thus, NAM states that auditory perceptual errors are a function of the intelligibility of stimulus words (*SWP*), the frequency of occurrence of stimulus words ($Freq_i$) and the similarity (*NWP*) and frequency ($Freq_j$) of neighboring words. The main hypothesis derived from NAM, and the central hypothesis to be tested in our experiments is that perceptual errors will increase as *p(ID)* decreases. Other things being equal, *p(ID)* will increase as the number, similarity, and frequency of neighbors decreases. There are, however, gaps between NAM-based work and the experiments we propose to do. First, the work on NAM has been done with one- and two-syllable English words [11, 17, 18]. In contrast, the average trademark drug name in the U.S. has 3 syllables and the average generic name has 4 syllables [19]. Second, *SWP* and *NWP* have been computed based on phonetic transcriptions of each stimulus word and empirically-gathered phoneme-to-phoneme confusion matrices, neither of which is currently available for drug names. Third, frequencies in NAM are based on the printed frequency of occurrence of various English words, but printed frequencies for drug names are neither available nor appropriate. Since we cannot follow the same procedures as Luce followed in computing *p(ID)*, a major focus of our work has been to make the NAM applicable to drug name confusions by developing and testing alternative methods for estimating *SWP, NWP* and $Freq_{i,j}$.

3. OPERATIONALIZING KEY QUANTITIES FROM THE MODEL

A. Phonemic Representation of Drug Names

Previous work using NAM used ordinary monosyllabic English words. Standard phonemic transcriptions for these words were extracted from electronic dictionaries. Although there are some references which contain drug name pronunciation guides, they are either in non-standard formats [20] or were incomplete and unavailable to us in electronic form [21]. Therefore we recorded pronunciations of roughly 6000 brand and generic drug names from one practicing nurse and one practicing pharmacist. These recordings were then transcribed to ARPAbet [22] by an experienced phonetician. Subsequent analyses were based on these ARPAbet transcriptions.

B. Confusion Matrices

Computation of both SWP and NWP requires reference to some source of phoneme-to-phoneme confusion matrices. The confusion matrices describe, for any given pair of phonemes, what the probability of confusion is. Element *i,j* of the matrix gives the probability that one will hear phoneme *i* when presented with phoneme *j*. We recognize that the confusion matrices we used were not ideal for our purposes. They were based only on single-syllable consonant-vowel-consonant words. The original confusion data were obtained at three different signal-to-noise ratios, where the noise was white noise. Confusion matrices (e.g., one at each signal to noise ratio, one for the beginning of syllables one for the end of syllables, etc.) were obtained from a separate project. We combined these into one composite matrix by taking the mean of all the component matrices.

C. Stimulus Word Probability (SWP)

Using the confusion matrices, computation of SWP was a straightforward application of Equation (2).

D. Neighborhood Word Probability (NWP)

Computing NWP was a two step process. Since a stimulus word and its neighbor may contain a different number of phonemes, the first step was to align the two phonemic strings. This was done using dynamic programming to compute the alignment between the two phoneme strings which minimized their phonological feature-based edit distance [23, 24]. Once the two words were optimally aligned, the phoneme-to-phoneme confusion probabilities were multiplied as shown in Equation (3). When one word was longer than another, it was necessary to have a confusion probability with the null phoneme. Luce's matrices contained this information. These probabilities corresponded to the situation when a subject was presented with a phoneme but produced no response or when the subject was presented with a null stimulus and produced a phoneme in response.

We are currently experimenting with a more elegant method which uses dynamic programming to align the two phoneme strings so as to maximize the sum of phoneme-to-phoneme log probabilities of confusion, rather than minimizing a sum of their edit distances. Since these phoneme-to-phoneme probabilities are given by the composite confusion matrix taken directly from Luce's tables, the antilog of the "distance" returned by this alignment function is precisely the required NWP, and no second stage of processing is required.

E. Frequency

Word frequency exerts powerful effects on word perception [11, 25, 26]. In psycholinguistics, frequency information is normally extracted from databases of printed word frequencies (or, in more recent times, from large electronic collections of text and speech) [27, 28]. Data on the printed frequency of drug names was both unavailable and not directly relevant. Instead, we used prescribing frequency data from several public and proprietary sources [29-32]. These data themselves are quite complex [33]. We used the average frequency from these 5 data sets to compute the frequency weights in equations (4).

F. Frequency-Weighted Neighborhood Probability (FWNP)

Using the methods described above, we computed FWNP for 5864 one-word drug names. The distribution of FWNP is given in Figure 3. Although Equation (4) indicates that FWNP should be based on all of the words in the lexicon, our pre-testing showed that using more than the 50 closest neighboring words had little effect. The scores in Figure 3 are based on the 50 closest neighbors of each drug name, where similarity was based on phoneme edit distance. In this figure, the horizontal axis represents the estimated probability of correct identification (i.e., FWNP). This result shows what one would expect from a carefully engineered lexicon like the drug name lexicon, namely, that most names have high probabilities of correct identification, with relatively few near neighbors. But the distribution is clearly bimodal. There are a disturbingly large number of names with very low FWNP scores. These names exist in dense, high frequency neighborhoods. Some of this may be due to the presence in the database of spelling variants of the same name (e.g., *Rephresh* and *Refresh*, *Carrasyn* and *Carrisyn*), but in other cases, we simply observe crowded, confusing neighborhoods (e.g., *Vivox*, *Zyvox*, *Vioxx*).

4. EXPERIMENTS

A. Overview of Experimental Design

We designed an experiment to examine the effects of frequency-weighted similarity neighborhoods on auditory perceptual accuracy. We chose the perceptual identification task for three main reasons: (1) it has ecological validity–in the real world health professionals and lay people must correctly identify spoken drug names in noisy environments; (2) the main outcome measure in perceptual identification is a categorical, correct/incorrect score, rather than simply a

reaction time; and (3) performance on the perceptual identification task is known to be sensitive to differences in the frequency and similarity neighborhood characteristics of words. The first

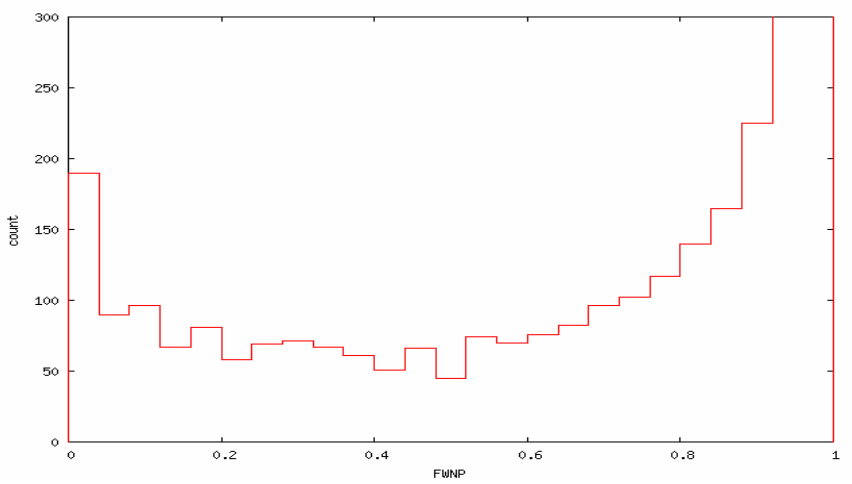

Figure 3: Distribution of FWNP scores for 5864 one-word drug names.

consideration was to select a task that was a valid simulation of what health professionals and patients do. Physicians, pharmacists, nurses, and lay people must often identify spoken drug names in noisy environments (e.g., retail pharmacies, emergency rooms, operating rooms, busy inpatient wards) and/or through noisy, narrow-bandwidth communication channels (e.g., telephone). Hence, the noise-masked, bandwidth-limited perceptual identification task seemed a logical choice.

B. Selection and Preparation of Stimulus Drug Names

As stimuli for the main auditory perception experiment, we selected one hundred brand and one hundred generic drug names. Names were stratified by prescribing frequency by taking roughly ten of each type from each decile of prescribing frequency. The names were then digitally recorded in a professional recording studio. They were read by a woman who was a trained phonetician and also an experienced voice-over actor. Pronunciations were based on the phonemic transcriptions described in Section 3.A. Each recorded pronunciation was then edited into a separate AIFF audio file. Since our perceptual identification experiments were intended to mimic the bandwidth limitations of telephone audio, frequencies below 300Hz and above 3000Hz were filtered out [34].

C. Background Noise

Drug names were played against a background of standard 20-speaker babble (obtained from Auditec of St. Louis). The noise was played at 65dB and was not bandwidth limited. The stimuli were played at either 63dB, 68dB, or 73dB, resulting in three signal-to-noise conditions of -2dB, +3dB, and +8dB respectively.

D. Procedure

Participants began by completing a demographic questionnaire. They then took a pure tone hearing threshold test. Participants whose hearing was below normal were excluded. Participants were then seated at a Macintosh PowerBook computer and fitted with a pair of headphones with an attached headset microphone (Beyerdynamic BT190). The participant then read the instructions. Playback of the 20-speaker babble was initiated, and this noise continued throughout the duration of the experiment. The PsyScope experiment program was used to run

the main experiment. The task began with 21 practice trials and then continued with 190 trials comprising the main experiment. Ten names from the original 200 were dropped to ensure the entire experiment took no more than one hour to complete. On each trial, the participant was asked simply to repeat back the name they think they heard. Spoken responses were captured through the headset microphone and digitally recorded on the computer. Reaction times (i.e., time between end of word playback and onset of response) were also captured by PsyScope. After completing the main experiment, participants moved to a different computer, put on a headset, and then pronounced the 190 experimental names as they were visually presented on the computer screen. After reading and pronouncing each name, participants rated their subjective familiarity with the name on a 7-point semantic differential scale. This final part of the task allowed us to capture pronunciation variation and subjective familiarity for later analysis.

E. Scoring

Recorded responses will be transcribed to ARPAbet. These transcriptions will then be compared to the ARPAbet transcriptions used to generate the recorded stimuli. We plan to employ both strict and lenient scoring criteria. The strict criterion will require complete, exact matching between response and reference transcriptions. Lenient criteria are still under development, but will be designed to capture legitimate pronunciation variation.

5. DISCUSSION

This paper has described work in progress. We have collected data from 66 pharmacists, and we plan to collect data from 50 physicians, 50 nurses, and 50 lay people in the coming year. Several significant challenges remain. Perhaps the most significant concerns how we will deal with pronunciation variation both in the computation of FWNP and in the scoring of responses. In addition, we have concerns about the validity and appropriateness of the confusion matrices we used to compute FWNP since the noise used to collect those matrices differed from the noise in our task. Similarly, we had to average across several disparate sources of frequency estimates to arrive at frequencies for each drug name. All of these concerns raise the possibility of measurement error, and at present we do not have good methods for quantifying its magnitude. There are also concerns about the ecological validity of our task. Although many of the pharmacists said that the task resembled their everyday struggles to correctly identify drug names, they also said the task was quite difficult and the noise was somewhat different from the noise in their work environments. In spite of these concerns, we believe our work has made significant progress on the problem of drug name confusion both theoretically and methodologically, and we hope this progress will improve patient safety.

ACKNOWLEDGEMENTS

This work was supported by grant number 5 R01 HS011609 from the U. S. Agency for Healthcare Research and Quality. We are grateful to Robert Gibbons, David Lewis, and David Tcheng, Clement Yu and the American Pharmacists Association for their assistance.

REFERENCES

1. L. T. Kohn, J. Corrigan and M. S. Donaldson, (Editors), *To err is human: building a safer health system,* (Institute of Medicine, Washington, DC, 2000).
2. M. Cohen, (Editor *Medication errors,* (American Pharmaceutical Association, Washington, DC, 1999).
3. N. M. Davis, "Drug names that look and sound alike," Hospital Pharmacy, **32(11)**, 1558-1570 (1997).
4. U. S. Pharmacopeia, "Similar names," USP DI Update Volume I and II, 4-7 (1997).
5. C. C. Fraterrigo and G. M. Frangella, "The "Z" word," American Journal of Hospital Pharmacy, **51(January 15)**, 235 (1994).

6. B. L. Lambert, K. Y. Chang and S. J. Lin, "Effect of orthographic and phonological similarity on false recognition of drug names," Social Science & Medicine, **52**, 1843-1857 (2001).
7. B. L. Lambert, K.-Y. Chang and P. Gupta, "Effects of frequency and similarity neighborhoods on pharmacists' visual perception of drug names," Social Science & Medicine, **57**, 1939-1955 (2003).
8. B. L. Lambert, S.-J. Lin, S. K. Gandhi and K.-Y. Chang, "Similarity as a risk factor in drug name confusion errors: The look-alike (orthographic) and sound-alike (phonological) model," Medical Care, **37(12)**, 1214-1225 (1999).
9. E. A. Flynn, K. N. Barker and B. J. Carnahan, "National observational study of prescription dispensing accuracy and safety in 50 pharmacies," J Am Pharm Assoc (Wash), **43(2)**, 191-200 (2003).
10. Quality Interagency Coordination Task Force, "Doing what counts for patient safety: Federal actions to reduce medical errors and their impact," Quality Interagency Coordination Task Force, Washington, DC, (2000).
11. P. A. Luce and D. B. Pisoni, "Recognizing spoken words: The neighborhood activation model," Ear & Hearing, **19(1)**, 1-36 (1998).
12. J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," Cognitive Psychology, **18**, 1-86 (1986).
13. D. Norris, "Shortlist: a connectionist model of continuous speech recognition," Cognition, **52**, 189-234 (1994).
14. M. Coltheart, E. Davelaar, J. T. Jonasson and D. Besner, "Access to the internal lexicon," in *Attention and Performance VI*, edited by S. Dornic (Academic Press, New York, NY, 1977).
15. W. D. Marslen-Wilson, "Functional parallelism in spoken word recognition," Cognition, **25**, 71-102 (1987).
16. R. D. Luce, *Individual choice behavior* (Wiley, New York, 1959).
17. P. A. Luce, S. D. Goldinger, E. T. Auer, Jr. and M. S. Vitevitch, "Phonetic priming, neighborhood activation, and PARSYN," Perception and Psychophysics, **62**, 615-625 (2000).
18. P. A. Luce, D. B. Pisoni and S. D. Goldinger, "Similarity neighborhoods of spoken words," in *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, edited by G. T. M. Altmann (MIT Press, Cambridge, MA, 1990).
19. B. L. Lambert, K. Y. Chang and S. J. Lin, "Descriptive analysis of the drug name lexicon," Drug Information Journal, **35**, 163-172 (2001).
20. United States Adopted Names Council, *USP Dictionary of USAN and international drug names* (U. S. Pharmacopeia, Inc., Rockville, MD, 1998).
21. *Physician's Desk Reference* (Thomson Healthcare, Montvale, NJ, 2005).
22. D. Jurafsky and J. H. Martin, *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (Prentice Hall, Upper Saddle River, NJ, 2000).
23. W. M. Fisher and J. G. Fiscus, "Better alignment procedures for speech recognition evaluation," *Proc. International Conference on Acoustics, Speech and Signal Processing, ICASSP-93*, Minneapolis, MN, (Year) p^ 59-62.
24. W. M. Fisher, J. G. Fiscus, A. Martin, D. S. Pallett and M. A. Pryzbocki, "Further studies in phonological scoring," Proceedings of the Spoken Language Systems Technology Workshop, 181-186 (1995).
25. R. L. Solomon and L. Postman, "Frequency of usage as a determinant of recognition threshold for words," Journal of Experimental Psychology, **43**, 195-210 (1952).
26. H. B. Savin, "Word-frequency effect and errors in the perception of speech," Journal of the Acoustical Society of America, **35**, 200-206 (1963).
27. A. Kilgariff, "Putting frequencies in the dictionary," International Journal of Lexicography, **10(2)**, 135-155 (1997).
28. F. Kucera and W. Francis, *Computational analysis of present-day American English* (Brown University Press, Providence, RI, 1967).
29. IMS Health, "NPA Plus: National prescription audit," IMS Health, Inc., Westport, CT, (2003).
30. National Center for Health Statistics. Ambulatory Health Care Data: NAMCS Description. [Web page]. Available at: http://www.cdc.gov/nchs/about/major/ahcd/namcsdes.htm. Accessed May 31, 2001, 2001.
31. National Center for Health Statistics. Ambulatory Health Care Data: NHAMCS Description. [Web page]. Available at: http://www.cdc.gov/nchs/about/major/ahcd/nhamcsds.htm. Accessed May 31, 2001, 2001.
32. Solucient Inc. Hospital drug utilization database. [Web page]. Available at: http://www.solucient.com/solutions/Solucients_Databases.shtml. Accessed March 5, 2004.
33. S.-J. Lin, B. L. Lambert, H. Tan and S. W. Toh, "Frequency estimates from prescription drug datasets," Journal of Pharmacoepidemiology and Drug Safety, ((under review)).
34. J. Rodman. The effect of bandwidth on speech intelligibility. Available at: http://www.polycom.com/common/pw_cmp_updateDocKeywords/0,1687,1809,00.pdf. Accessed June 17, 2005.