# Determinants of Wordlikeness

**Todd M. Bailey (todd.bailey@psy.ox.ac.uk)**
Department of Experimental Psychology
University of Oxford
South Parks Rd.
Oxford OX1 3UD, U.K.

**Ulrike Hahn (U.Hahn@warwick.ac.uk)**
Department of Psychology
University of Warwick
Coventry CV4 7AL, U.K.

**Abstract**

Wordlikeness, which is generally equated with phonotactics, is becoming an increasingly important variable in the study of language acquisition and processing as well as in the context of verbal short term memory. Past research has sought to establish phonotactic knowledge (knowledge of the possible sequences of sounds within a language) as a distinct kind of knowledge above and beyond knowledge of individual lexical items. It is unclear, however, how separate phonotactic and lexical knowledge really are; conceivably there could be effects of similar sounding lexical neighbors on perceived wordlikeness. We report empirical evidence and analysis demonstrating independent contributions of phonotactics and of lexical neighbors in accounting for wordlikeness ratings, a finding with both methodological and cognitive implications.

## Introduction

Any speaker of English can tell that Zbigniew is not an English name; any such speaker knows it not only does not happen to be an English name but there is a strong intuition that it could not be one either. Polish words are clearly different from English words, and speakers can tell this intuitively even if they cannot articulate why. Furthermore, even among actual words of English, some words sound more "typical" than others, say *rat* as opposed to *sphere, splurge,* or *phlegm.*

Thus given a known or unknown sequence of sounds, we have intuitions on whether the sequence could be a word of English and how natural it would sound as a word of English. This sense of "wordlikeness" is the focus of this paper.

Wordlikeness is implicated as a relevant factor in a variety of domains, including speech perception, lexical development, and verbal memory, which establish knowledge of wordlikeness as an important type of linguistic knowledge. Research in these domains has made use of several conceptually related ways of measuring wordlikeness. Our motivation for the research reported here stems from our own attempts to use these measures for the principled construction of experimental stimuli, both words and nonce-words, as would be desirable in a wide variety of psycholinguistic domains. Contrary to our expectations, the use of standard measures of wordlikeness did not seem to correspond to our intuitive judgements of the resultant stimuli in any real way. This prompted us to examine more closely the determinants of wordlikeness.

Thus, our work has a methodological point of departure. Due to the widespread appeal to wordlikeness as a type of linguistic knowledge, however, the research described here is also a question about the nature of linguistic knowledge of intrinsic interest; and, as will become apparent, it is a question in which resurfaces one of the most ubiquitous issues concerning mental representation in Cognitive Science, namely the extent to which knowledge is abstract or merely implicit in stored exemplars.

### Measures of Wordlikeness

Wordlikeness, as evidenced in our intuitive judgements, is typically equated with *phonotactics.* Phonotactics is the term used in linguistics to refer to the sequential arrangements of phonological units (sounds) which occur in a given language. The term has two related senses. It can be used to refer to the *constraints* on what constitutes a permissible or acceptable sequence in the language and it can be used in a more fine-grained sense which refers to the *probability* with which particular sequences occur. Phonotactics as sequence acceptability highlights, e.g., the fact that there are no words in English which start with the consonant sequence /fs/ or /spm/, and that there seems to be a strong sense in which English could not have such words, thus indicating a phonotactic constraint. Phonotactics as sequence probability is concerned with the frequency with which a particular sequence occurs in the language and is generally based on measures or counts taken over large language samples. In this latter sense, phonotactics deals with *graded* measures of phonological well-formedness with the consequence that a sequence of sounds, whether an existing word or a novel sequence, can be more or less wordlike, more or less typical of English words.

## The Role of Wordlikeness

Wordlikeness has been established as a relevant type of linguistic knowledge in an increasing variety of domains.

The first of these is spoken word perception. Brown and Hildum (1956) found that participants asked to identify monosyllabic sequences embedded in noise were better at identifying real English words than at identifying non-words, but they were also better at identifying non-word sequences which are phonotactically permissible in English compared to sequences which are illegal. Thus they showed both knowledge and effects of phonotactic constraints. Another study, by Vitevitch, Luce, Charles-Luce and Kemmerer (1997), found effects of phonotactic probability on participants' reaction times in an auditory repetition task: high-probability nonwords were responded to more quickly than low probability nonwords.

A second area where wordlikeness has been shown to be relevant is developmental research which has revealed infant sensitivity both to phonotactic constraints and to sequence probabilities. Jusczyk, Frederici, Wessels, Svenkerud, and Jusczyk (1993) found that 9 month old infants listened longer to lists of words in their native language, preferring them over foreign words which violated native phonotactic constraints. Jusczyk, Luce and Charles-Luce (1994) found that 9 month old infants preferred to listen to lists of monosyllabic nonsense words with high probability phonotactic sequences (in English) over lists containing low probability sequences.

These results link suggestively with recent computational work demonstrating the utility of phonotactic constraints (Brent and Cartwright, 1996) and phonotactic probabilities (Cairns, Shillcock, Chater and Levy, 1997) for segmentation of the speech stream. Thus phonotactics might provide an important cue which the infant can use to bootstrap lexical acquisition.

A third area in which wordlikeness has figured prominently is verbal short term memory. It has recently been demonstrated that verbal short-term memory interacts with long-term knowledge, rather than being an independent buffer. One source of evidence in this context are studies on the so-called "lexicality effect", the robust finding that participants are better at recalling lists of familiar words than lists of the same length containing English-sounding nonwords (Hulme, Maughan and Brown, 1991). Subsequently, very recent research in this area has directly demonstrated a "wordlikeness effect" in nonword repetition—a nonword analogue to the lexicality effect according to which greater wordlikeness supports greater repetition accuracy in both children (Gathercole, Willis, Emslie and Baddeley, 1991; Gathercole, 1995) and adults (reported in Gathercole and Martin, 1997). For measures of wordlikeness, these studies have relied both on explicit wordlikeness judgements, where participants were asked to rate wordlikeness on a scale (Gathercole et al., 1991), and on phonotactic probabilities (Gathercole, 1995).

Additional interest in this work stems from the emerging links between verbal short-term memory and vocabulary acquisition. Childrens' scores on tests of verbal short-term memory —such as nonword repetition— are predictors of the ease with which they acquire novel words (Gathercole and Baddeley, 1989; Gathercole, Willis, Emslie and Baddeley, 1992; Gathercole, Hitch, Service and Martin, 1997) and of their ability for foreign language learning (Service, 1992). There is also some evidence that children with developmental language disorders have poorer vocabulary knowledge due to impaired phonological memory skills (Gathercole and Baddeley, 1990). In short, wordlikeness is becoming an increasingly important variable in explanations of cognitive performance, with research expanding particularly rapidly in the context of language development.

## The Nature of Wordlikeness

All of the domains in which wordlikeness has been implicated require some way of measuring wordlikeness. Thus all immediately raise the question of what constitutes knowledge of wordlikeness, what exactly are the factors that influence it. Though all the research cited has had *some* means of measuring wordlikeness, and these measures have proved useful in analyzing behavioral data, our own experience, mentioned above, suggests that these issues might not be resolved.

What factors determine wordlikeness, i.e., what type of knowledge wordlikeness implicates, is the first and most fundamental question to ask about wordlikeness. But there is a further question involved in understanding the nature of this type of knowledge, namely the nature of the mental representations underlying it. Specifically, the question is whether this knowledge is merely *implicit* in our lexical knowledge, that is within the individual words we know, or whether the relevant information is extracted from the lexical level and stored separately (Vitevitch, 1997), an issue which has been given attention in the context of models of speech perception (Norris, 1994) and memory (Gathercole and Martin, 1997).

The research presented in this paper concentrates on the first question, the factors determining wordlikeness (and the resultant implications for measurement), but as will become apparent, the particular factors we are investigating also make this work directly relevant to the representation issue, in particular with respect to the lexicon.

Before we describe our own study, however, a closer look at previous work on the factors governing wordlikeness is necessary.

# Previous Work

Though there has been a sizeable amount of work making use of wordlikeness, there have been virtually *no direct* investigations of the factors determining wordlikeness.

Some of the studies described above, e.g., Gathercole et al., (1991), simply use wordlikeness judgments as a measure of wordlikeness. Here, participants are required to rate words (or nonwords) on a numerical scale reflecting "goodness", or "typicality" or "the degree to which this word sounds like a word of English". Mean wordlikeness rating is then subsequently used as a predictor.

Those studies which employ phonotactic measures instead of wordlikeness ratings at least offer *indirect* information on the factors governing wordlikeness. To the extent that the calculated phonotactic measure successfully predicts behavior, this success provides evidence for the relevance and efficacy of the calculated factors. The most frequent measure of phonotactic probability by far is simple bigram statistics, that is a measure of how frequently any two phonemes co-occur (e.g., how probable the occurrence of /o/ after /t/ is). Similarly, trigram statistics could be computed (e.g., Cairns et al., 1997) or measures which more closely take into account syllable structure. Bigram co-occurrence probabilities, or weighted variants thereof, were used for instance in Jusczyk et al. (1994), Gathercole (1995).

A more direct investigation of wordlikeness was conducted by Vitevitch et al. (1997). They used (log-frequency weighted) bigram co-occurrence probabilities and positional segment frequencies (that is, how often a given phonetic segment occurs in a particular position of a word) to generate non-words of high and low phonotactic probability. These were then used in a wordlikeness rating task, which yielded significantly higher mean ratings for high-probability than for low probability non-words, thus confirming the relevance of bigram statistics for wordlikeness. Their second study demonstrated with the same non-word set that high-probability words were responded to more rapidly than low-probability words in a speeded repetition task. Thus, the two studies provide support for bigrams as a relevant factor both through a judgement task and a behavioral measure. However, this work (which also sought to examine the effects of syllable stress) did not seek to quantify directly how large the contribution of phonotactics was.

While phonotactics has been the dominant way to conceptualize wordlikeness, there are two studies which have taken the lexicon as their point of departure. Greenberg and Jenkins (1964) constructed nonwords which varied in "distance" to real words under a transformation-based measure, i.e., the number of phoneme substitutions required to transform the nonword into an actual word of English. Some nonwords were constructed entirely of phonemes of English, whereas others also included non-native sounds. Greenberg and Jenkins (1964) found high correlations between their transformation-based measure and participants' judgments of how close items were to "being a word of English".

This emphasis on the lexicon is continued in the most direct investigation of wordlikeness to date is, a study by Martin and Gathercole, of which a preliminary report is given in Gathercole and Martin (1997). This study took wordlikeness ratings of spoken nonwords and directly investigated the impact of a hitherto unconsidered factor, neighborhood density, on ratings. Neighborhood density measures were derived from lists of lexical neighbors generated by participants. Specifically, for each nonword, participants were asked to produce all the real words which the nonword "brought to mind", thus yielding several different subjective measures of neighborhood and neighborhood density. The best of these measures accounted for 21% of the unique variance in the ratings, thus indicating that lexical density of the phonological space occupied by a word or nonword is an important contributing factor. So far, however, no information about phonotactic probabilities has been reported on this project. Thus, like Greenberg and Jenkins (1964) it does not link in with the phonotactic based research.

Thus extant research leaves a lot unanswered. A considerable variety of measures have been used, both phonotactic and lexically based in nature, thus raising the question of which are the most appropriate. This is a question which should be pursued *directly*, not just by demonstrating that a particular measure such as bigrams can be used to predict significant differences in performance, but by directly estimating the size of that measures' contribution to performance. There is good evidence that bigram statistics have an impact on wordlikeness, but given our own experiences bigram statistics might be only one, comparatively small, factor influencing perceived wordlikeness.

Furthermore, pursuing the comparatively ignored issue of the potential impact of lexical knowledge seems vital. This follows also from the general debate about the nature of the underlying mental representations as discussed above. Wordlikeness knowledge might simply be implicit in our lexical knowledge, or it might be knowledge which is compiled and extracted from the lexicon, as are phonotactics. Deciding on the representational nature of this knowledge thus requires a careful search for both phonotactic and lexical contributions.

## Experimental Investigation

The study elicited ratings of wordlikeness for a set of nonwords and a set of words, randomly intermixed. The aim of the study was to assess the utility of various phonotactic measures as well as to test whether and how wordlikeness is affected by the phonological distance between a probe item (word or nonword) and nearby words, and also to test whether the frequency of these nearby words affects wordlikeness judgments.

### Participants

Participants were 10 first year psychology students from Warwick University who took part in the experiment as part of their Methods course.

### Materials

Stimuli were selected by first extracting monosyllables from the CELEX database[1] of English words (these monosyllables are henceforth referred to as the set of English words). The use of monosyllabic stimuli avoided the added complexity of stress placement and syllabification. A syllable formation grammar was constructed based on the set of all onset-nucleus and nucleus-coda combinations observed in the English (monosyllabic) words, and the grammar was used to

[1] CELEX can be obtained by contacting celex@mpi.ne

generate all "pronounceable" word forms, i.e. phonological syllables consistent with onset-nucleus and nucleus-coda transitions observed in English monosyllables. One or more nearest neighbors was identified for each word form by comparing it with the set of English words. Each word form was classified as an *isolate, near miss, or word.* Word forms which differed from their nearest real word neighbor by two phonemes were classified as *isolates.* Word forms which differed from the nearest neighbor by a single phoneme were classified as *near misses.* From the *isolates*, a subset of 22 items were chosen. These 22 *isolates* were included in the stimulus set, along with 271 *near misses* which were one phoneme different from any of the 22 *isolates.*

## Procedure

Words were presented in a written questionnaire. Participants were instructed to judge the sounds of the words, not their spellings, and to indicate the perceived typicality of the word or nonword by circling the appropriate number on a scale from 1 to 9. For example,

Does minth sound like a typical word of English?

Very non-typical = 1 2 3 4 5 6 7 8 9 = Very typical

Real words and nonwords were intermixed, so as to heighten the naturalness of the nonword rating task.

## Results

**Phonological Well-Formedness** Three measures of phonological well-formedness were computed for each probe item based on monosyllables in the CELEX English database. This set contains 7239 phonologically distinct entries, i.e. different syllables, among the 24,859 monosyllabic entries, including homophones, in total.

The phonological well-formedness measures computed were bigram, trigram, and syllable part transition probabilities. The bigram probability for each item was computed by calculating the conditional probability of each phonological segment in the item, given the preceding segment. This involves counting the number of times, say, /t/ occurs after /o/, etc., in the monosyllabic subset of the CELEX lexicon. Word boundaries counted as segments in this analysis. The geometric mean of conditional segment probabilities was computed to arrive at a single average bigram probability for the entire item. Trigram probabilities were computed in a similar fashion by calculating the conditional probability of each segment given the preceding two segments. This involves counting the number of times /t/ occurs after /po/, or after /es/, etc. A syllable part probability was computed for each item taking syllable onset, nucleus and coda to be the basic units of analysis rather then segments. The syllable part probability for an item was computed from the geometric mean of the probabilities of its onset-plus-nucleus and its nucleus-plus-coda.

Bigram and trigram probabilities were computed with and without weighting based on the log token frequency given for monosyllabic words in the CELEX lexicon. Corresponding measures with and without token frequency weighting were highly correlated ($r = .98$ for bigrams, $r = .99$ for trigrams) and inspection revealed no clear non-linear relationships between weighted and unweighted variables. Subsequent calculations were based on probabilities without frequency weighting. Bivariate correlations between bigram, trigram, and syllable part probabilities were 0.38 (bigram/trigram), 0.44 (bigram/syllable part), and 0.40 (trigram/syllable part).

| Enhanced Model | $R^2$ | Fchange (df) |
|---|---|---|
| SYL PART + NNB | 0.12 | F(1,256) = 34** |
| BIGRAM + NNB | 0.11 | F(1,256) = 33** |
| TRIGRAM + NNB | 0.17 | F(1,256) = 51** |
| SYL PART + BIGRAM + TRIGRAM + NNB | 0.17 | F(1,254) = 52** |

Table 1: Table of results indicating the increases in proportion of the variance in ratings explained by adding neighborhood size to the model. Fchange is an F statistic based on the additional variance accounted for in the enhanced model (taking neighborhood size into account) compared to a simple model including only phonological well-formedness, ** denotes significance levels of $p < .0001$.

**Data fits** To assess the extent to which wordlikeness ratings might be determined by simple phonotactic probabilities, linear regression models were fit to the nonword ratings which were averaged across subjects to obtain a single mean wordlikeness rating for each probe. All three measures of phonotactic well-formedness were significantly correlated with ratings, though none of them accounted for more than about 7% of the variance between items ($R^2 = .07, p < .0001$; $R^2 = .05, p = .0002$; $R^2 = .02, p = .03$; for syllable part, trigram and bigram probabilities, respectively). If all three measures of well-formedness are included, the combined model accounts for about 9% of the variance ($R^2 = .09, p < .0001$).

As a first and very crude test for neighborhood effects, each of the above regression models was enhanced by adding a variable representing the neighborhood size for each probe item, that is, the number of real words which differ from the probe item by no more than two phonemes. In each case, neighborhood size improved the fit of the model significantly (see Table 1).

This suggests that wordlikeness ratings were at least partially influenced by their phonological neighborhood, that is, by the presence in the lexicon of words of similar pronunciation. Disembodied phonotactic well-formedness alone is not sufficient to account for the ratings. In fact, even as crude a measure as neighborhood size accounts for 9% of the variance on its own. This is of the same magnitude as the contribution of phonotactics, but interestingly it is not shared variance: both neighborhood size and phonotactic measures account for a largely different 9% of the variance as is apparent from the fact that, together, they account for 17%.

We also tested a slightly more refined measure of neighbors

and neighborhoods, a variant of the neighborhood size measure, where each neighbor is weighted by proximity to the probe, as measured by the number of shared syllable parts. This model is based on the large psychological literature that has systematically related generalization to distance in an internal psychological space for a variety of tasks such as identification, old-new recognition, or categorization (Shepard, 1987; Nosofsky, 1988) and, most recently, inflectional morphology (Nakisa and Hahn, 1996) and artificial grammar learning (Pothos and Bailey, 1997). We (crudely) computed phonological distance between the probe and all other items by counting the number of syllable positions on which any two sequences differed.

This distance was then used to derive a composite similarity measure, $sim[i, LEX]$, between the probe and all other items in the (monosyllabic) lexicon.

$$sim[i, LEX] = \sum_j exp(-S * d[i, j])$$

where $i$ is the probe, $j$ is a word in the lexicon, and $d[i, j]$ is the (syllable part) distance between $i$ and $j$ and $S$ a sensitivity parameter that governs the rate of the exponential decay.

Even this crude measure of similarity on its own accounts for 13% of the variance, thus surpassing the combined effects of phonotactics, and further emphasizing the role of the lexicon.

### Discussion

Our first main result is the fact that peoples' intuitions about wordlikeness are not well captured by standard phonotactic measures alone. Though bigrams, trigrams and syllable-part transition probabilities all seem to have some influence, their individual contribution is small, and their joint contribution limited. Surprising too, was that weighting according to token frequency yielded such extremely highly correlations with non-weighted measures.

Our second main result is that our intuitions about wordlikeness are significantly affected by the lexical neighborhood of an item in phonological space. The degree to which the lexicon contains words which sound similar to an item affects its perceived wordlikeness. This result confirms those reported in Gathercole and Martin (1997). The analyses we have performed so far also suggest that neighborhood and phonotactics each make independent contributions. This result goes beyond the work reported by Gathercole and Martin (1997) by directly estimating independent contributions of lexicon and phonotactics.

As they stand our results suggest quite simply that the source of a particular type of knowledge of language we have, namely the extent to which a sequence forms a better or a worse instance of the words of a language is still poorly understood. This also suggests that there might be potential improvements to cognitive accounts and models in a variety of areas through incorporation of a notion of wordlikeness, which is richer than just phonotactics and which incorporates neighbor- and neighborhood effects as well.

It is, of course, possible that in some or even all of the domains in which wordlikeness is implicated, explicit judgments aside, phonotactics proper will turn out to be doing all the work. This does not, however, strike us as likely, given the rather limited contribution to intuitive judgments displayed here. The ultimate test, however, must be more detailed modelling work in these areas.

Our results are also informative with respect to the representational nature of wordlikeness knowledge in that they seem consistent neither with the idea that all knowledge of wordlikeness is given implicitly in the lexicon, nor with the idea that it is based entirely on a separate non-lexical source. There is an important caveat here. It remains to be seen whether more fine-grained models —either of phonological neighborhood or of phonotactics— might not be found which are sufficiently powerful to explain single-handedly what now appears to be a joint contribution. Candidate models for phonotactics could be statistics which include more long-distance or higher-order distributional properties; candidate models of neighbor effects could be models which incorporate the (token) frequency of the neighbors. Given the ubiquity of frequency effects in lexical processing (see e.g., Luce, Pisoni, and Goldinger, 1990), the latter will be our own next step.

## Conclusions

We have shown that intuitive judgements about wordlikeness, that is our intuitions about how typical a sound sequence is of words of the language, are not well-captured by phonotactics alone. They are influenced considerably by the lexicon — sequences have high wordlikeness partially because they are similar to other known words. This suggests that not only is a closer look at the knowledge underpinning these intuitions required, but also that the wide variety of domains in which either wordlikeness ratings or phonotactics have been implicated might profitably be revisited.

## Acknowledgements

Todd Bailey was supported by a McDonnell-Pew Fellowship in cognitive neuroscience.

## References

Brent, M.R. and Cartwright, T.A. (1996) Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61, 93-125.

Brown, R.W. and Hildum, D.C. (1956) Expectancy and perception of syllables. *Language*, 32, 411-419.

Cairns, P., Shillcock, R., Chater, N., and Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33, 111-153.

Gathercole, S.E. (1995) Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 83-94.

Gathercole, S.E. and Baddeley, A.D. (1989) Phonological memory deficits in language-disordered children: Is there a causal connection? *Journal of Memory and Language*, 28, 200-213.

Gathercole, S.E. and Martin, A.J. (1997) Interactive processes in phonological memory. In, Conway, M.A. (ed) *Cognitive Models of Memory*. Hove, U.K.: Psychology Press/ MIT Press.

Gathercole, S.E. and Hitch, G.J. and Service, E.S. and Martin, A.J. (1997) Phonological short-term memory and new word learning in children. *Developmental Psychology*, 33, 966-979.

Gathercole, S.E., Willis, C.S., Baddeley, A., Emslie, H. (1994) The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, 2, 103-127.

Gathercole, S.E., Willis, C.S., Emslie, H. and Baddeley, A. (1991) The influences of number of syllables and wordlikeness on childrens' repetition of nonwords. *Applied Psycholinguistics*, 12, 349-367.

Gathercole, S.E., Willis, C.S., Emslie, H. and Baddeley, A. (1992) Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887-898.

Greenberg, J.H. and Jenkins, J.J. (1964) Studies in the psychological correlates of the sound system of American English. *Word*, 20, 157-177.

Hulme, C., Maughan, S. and Brown, G.D.A. (1991) Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30, 685-701.

Jusczyk, P.W., Frederici, A.D., Wessels, J., Svenkerud, V.Y. and Jusczyk, A.M. (1993) Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, 32, 402-420.

Jusczyk, P.W., Luce, P.A. and Charles-Luce, J. (1994) Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.

Luce, P.A., Pisoni, D.B., and Goldinger, S.D. (1990) Similarity neighborhoods of spoken words. In, Altman, G.T.M. (Ed.) *Cognitive Models of Speech Processing: Psycholinguistic & Computational Perspectives*. MIT Press.

Nakisa, R.C and Hahn, U. (1996) Where defaults don't help: the case of the German plural system. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Nosofsky, R.M. (1988) Exemplar-based accounts of the relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 700-708.

Pothos, E.M and Bailey, T.M. (1997) Rules vs. Similarity in Artificial Grammar Learning. In *Proceedings of SimCat97, an Interdisciplinary Workshop on Similarity and Categorization*. Dept. of Artificial Intelligence, University of Edinburgh.

Shepard, R.N. (1987) Toward a Universal Law of Generalization for Psychological Science, *Science*, 237, 1317-1323.

Service, E. (1992) Phonology, working memory and foreign language learning. *Quarterly Journal of Experimental Psychology*, 45A, 21-50.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J. and Kemmerer, D. (1997) Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40, 47-62.