# Constructing Typing-Time Corpora: A New Way to Answer Old Questions

**Uriel Cohen Priva (urielc@stanford.edu)**
Department of Linguistics, 450 Serra Mall
Stanford, CA 94305 USA

**Abstract**

Many current studies in linguistics and psycholinguistics require the use of phonetically labeled speech data. Collecting and annotating such data is expensive and slow. An alternative approach makes use of pre-labeled speech corpora, but these are available for very few languages, might not contain the desired linguistic environment, and the construction of new ones is still expensive and time-consuming. We present a fast and cost-efficient method for constructing a new type of corpus which retains many of the advantages of phonetically labeled speech, *typing-time corpora*. In this paper we show that an English typing-time corpus collected over the web is sufficient to replicate word frequency and neighborhood density effects. We then demonstrate the transferability of this method to less studied languages and to different orthographies. We show that a smaller Hebrew typing corpus collected over the web can be used to find lengthening effects in infrequent Hebrew words.

**Keywords:** Typing-time; Corpora; Frequency; Neighborhood density; Amazon Mechanical Turk

## Introduction

Many studies in linguistics and psycholinguistic require either the precise annotation of durations and latencies for speech data gathered in carefully controlled experiments, or the availability of phonetically labeled corpora. For example, evidence for *neighborhood density* in production depends on measuring speech production latencies (Vitevitch, 2002). The study of *production difficulties* relies on measuring the lengthening of words in a difficult context (Fox Tree & Clark, 1997). Studies of *frequency* and predictability-dependent phonetic reduction (Van Son & Van Santen, 2005; Pluymaekers, Ernestus, & Baayen, 2005; Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009) make use of corpora containing exact word and phone durations, such as the Switchboard Corpus (Godfrey & Holliman, 1997) and Buckeye Corpus of Conversational Speech (Pitt et al., 2007) for English, the Spoken Dutch Corpus (Oostdijk, 2000) and the Kiel Corpus of Spontaneous Speech (Kohler, Pätzold, & Simpson, 1995) for German.

However, neither the experimental approach nor the corpus-based one may be a feasible option when trying to address the problem of data availability in less studied languages. Subjects may not always be available on the one hand, and Switchboard-like corpora do not exist for most languages on the other hand. In addition, even when a corpus is accessible, it might not contain the relevant linguistic environments for addressing the questions at hand. The creation of even a small-scale corpus is an expensive and time-consuming project, and therefore, there is much gain in finding a simpler alternative. In this paper, we propose a solution to this problem. We show that by wedding two methodological advancements, tracking typing speed and collection of data over the web, we can create an alternative both to experiments which require phonetic labeling and to phonetically-labeled corpora, *typing-time corpora* — corpora of typed data in which each letter and word is annotated with the time it took to type.

A number of studies (Weingarten, Nottbusch, & Will, 2004; Zesiger, Orliaguet, Boë, & Mounoud, 1994 among others) demonstrate that typing is sensitive to language-based effects. Weingarten et al. (2004) show that typing is sensitive to phonological and morphological properties of the words being typed. Zesiger et al. (1994) show that actual words are typed faster than pseudo-words and that frequent words are typed faster than infrequent words. These effects demonstrate that even though a typing task is different from spoken speech production, it does exhibit linguistic effects that are normally associated with speech.

Not only does typing time provides a window to linguistic performance, but it also holds a big advantage, as it allows the automatic gathering of large amounts of data through the web. A simple way to utilize this is by using Amazon Mechanical Turk (AMT), a virtual work marketplace created by Amazon.com. On AMT, requesters can upload work requests in the form of HTML pages, which workers can access online. Several researchers in the natural language processing community (Callison-Burch, 2009; Colowick & Pool, 2007; Snow, O'Connor, Jurafsky, & Ng, 2008 among others) make use of AMT to construct corpora for which human-labeled data is not available, or to annotate new data sets. In this paper, we demonstrate that extending the use of AMT to the construction of typing-time corpora provides an easy and cost-efficient alternative to laboratory experiments and extant corpora. We show evidence that supports the applicability of this methodology by constructing a typing time corpus for English, and using it to replicate two well known effects on language production: word frequency (Bell et al., 2009) and neighborhood density (Coltheart, Davelaar, Jonasson, & Besner, 1977; Vitevitch, 2002; Adelman & Brown, 2007). We then extend these results to a less studied language. We show the effect of word frequency on typing time in a smaller Hebrew corpus, exemplifying that the paradigm holds even for relatively small typing-time corpora, and for different languages with varying orthographies.[1]

## Previous production studies

### Frequency effects

Much current work in linguistics stresses the importance of word-frequency in the minute modulations in the duration of words, morphemes, syllables and phones in various contexts. These durations are taken from corpora of spontaneous or read speech in which phone durations were hand-labeled by linguists. Bell et al. (2009) show that frequent English words tend to reduce more than infrequent words. Pluymaekers et al. (2005) show the reduction of Dutch morphemes in predictable contexts. Aylett and Turk (2006) show reduction in predictable English syllables. Van Son and Van Santen (2005) show that some contextually predictable consonants are more likely to reduce.

### Neighborhood effects

A wide range of studies has shown language users to be sensitive to the effects of neighborhood-density (Coltheart et al., 1977; Vitevitch, 2002; Adelman & Brown, 2007; Peereman & Content, 1997). Coltheart et al. (1977) defines the neighborhood density of a given word as "the number of words that can be produced by changing just one of the letters in the string to another letter, preserving letter positions." Two different definitions of neighborhood density follow naturally from this one: Coltheart's original spelling-based definition, in which the substitutions are of single orthographic characters, and a phonological definition, in which the substitution is based on phonemes. Peereman and Content (1997) argue that the best approximation of neighborhood density is phonographic, that is, the cases in which the spelling neighbor is also the phonological neighbor. Furthermore, neighborhood density has been shown to have different consequences in production and comprehension. For English, Vitevitch (2002) shows that a dense neighborhood facilitates spoken word production, whereas Vitevitch and Luce (1998) show that a dense neighborhood inhibits word comprehension.

## Motivating typing-time corpora

The studies cited above demonstrate the benefit of investigating slight modulations of durations and latencies in spoken language production. However, many of them presuppose rather ideal experimental settings: a laboratory with accurate recording equipments, access to relevant human subjects in the proximity of that laboratory, sufficient time to label large amounts of data, and ample funds. The often easier alternative of using a pre-labeled speech corpus is not available when the linguistic environment being studied is not present in the corpus, or when no such corpus is present, as is in fact the case with most languages. Therefore, there would be much to benefit from a new methodology for investigating language production.

The following two sections describe the components of the solution we propose for this problem: an experimental approach to the collection of typing-time data, and the collection of large amounts of data over the web. By combining these methodologies we can create *typing-time corpora*, which provide an answer to the problem we presented above; they do not require any special equipment, subjects from remote locations can provide experimental data over the web, and no further labeling is required.

### Online data collection

Even basic web technology allows the collection of data through the web. Every search request on the web involves sending data to some webserver, which can collect the data it receives. However, utilizing web technology for data collection requires finding enough workers to perform the specific task. Amazon Mechanical Turk (AMT) provides a simple platform to do so. AMT is a virtual marketplace in which requests and workers can interact. The requester uploads tasks in the form of HTML pages to the website and proposes to pay a given price for the completion of each task. Workers can choose among available tasks, perform them, and submit the results through AMT. The requester can then review and approve the results, which leads to the transfer of the proposed sum of money from his account to the workers' accounts. AMT handles the overhead involving all other aspects of the interaction: the exchange of money and the collection of the results.

Several recent studies have already made use of AMT (Callison-Burch, 2009; Colowick & Pool, 2007). Colowick and Pool (2007) use AMT to find preferences for semantic scope ambiguity, and Callison-Burch (2009) uses AMT to evaluate the quality of automatic translations.

One possible concern with data collected this way is whether it can be as accurate as data collected under controlled conditions. However, Snow et al. (2008) compare the performance of AMT annotators with that of professional annotators, and they find that by increasing the number of annotators, untrained annotators over the web can match the performance of expert annotators. Increasing the number of data points per observation type is a key concept in handling noisy data collected

[1] The typing time approach is, of course, limited to languages that have a letter-based written standard (unlike, e.g., Chinese). While not all languages have such a written form, or any kind of written form at all, the proposed methodology would still allow access to a large number of currently less studied languages.

over the web. Since the gathering of data over the web is fast and inexpensive, enough data points can be collected to ensure that noisy data would be as sensitive as a smaller amount of data collected under ideal conditions.

### Typing time experiments

Several studies have demonstrated that typing speed is affected by linguistic factors. Gentner (1982) shows that a sequence of keystrokes is more predictive of the time it would take to strike one key if the sequence does not span word boundaries. Gentner, Larochelle, and Grudin (1988) show that the same four-key sequence is typed faster in frequent words than in infrequent words of comparable length. Weingarten et al. (2004) show that typing is sensitive to morphological-syllabic boundaries, by comparing the lag between typing two specific keys, held constant across conditions, and varying between syllable and morpheme boundaries.

Since typing requires moving the hands and fingers to different locations on the keyboard, the baseline lag between the typing of a given key and the preceding key varies dramatically based on the preceding and possibly the following keystrokes. Gentner (1982) shows that almost 50% of the variability is controlled for if we control for the immediately preceding keystroke. He also shows that adding up to one more key to the preceding context of the target key, and up to one following key, can account for most of the location-based variability.

While typing time studies clearly show the potential of using typing time as a segue to assessing linguistic performance, the factorial methods used in Gentner et al. (1988) and Weingarten et al. (2004) are not always replicable in further languages. Weingarten et al. (2004), who investigate lexical access effects in German, keep the same two-key sequences while varying the morphological and syllabic environment. However, many languages would not necessarily allow the same two-key sequence to appear in every condition, making a factorial design impossible. It would be beneficial to see such effects even if only some of the conditions exist for each two-key environment. Gentner et al. (1988) uses identical four-key sequences embedded in words of varying frequencies, but in orthographic systems in which vowels are not assigned a separate letter (e.g. Arabic or Hebrew), words that contain identical four-key sequences would usually belong to the same stem or the same neighborhood. These issues can be remedied by the proposed methodology of constructing a typing-time corpus.

## Building typing-time corpora

We construct the typing-time corpus in the following manner. AMT workers (or other web users) are presented with an HTML form in which they first fill in some basic details. We request our subjects to say whether they are left- or right-handed, and whether they look at the keyboard while typing. They are also requested to type in the keyboard keys below the digits 1–6 in order to identify the keyboard layout, and to fill in the first two languages they speak, following an example in which the first language is not the language we want to investigate. In order to reduce the variance, submissions from anyone who is left-handed, looks at the keyboard, is not using the most common keyboard layout (QWERTY in the case of English) or did not fill in the language we want to investigate were not included in the analysis (but were still accepted and paid).

After the basic details are collected, the subjects move to ten open text fields. After they choose the field, text appears to the right (and in right-to-left languages, to the left) of the open text field, and the subjects are instructed to copy it. Once they move to the next field, the field they leave is locked, and they are no longer able to change it. While they type, a javascript program running in the web page collects the exact time of each key press.

The output of the collected data is then parsed and assigned additional attributes. Each keystroke is associated with the word it belongs to and the key typed in that word. Corrected text is recorded as *corrected*, and words that contain it are marked as *corrected.* Words that do not match the target text are recorded as *wrong.* Keystrokes that took more than 500ms to type are considered a *break*, and words that contain breaks past the first characters are considered *interrupted.* When a word is marked as *interrupted*, *corrected* or *wrong*, all the keystrokes that comprise it are marked as having a *interrupted*, *corrected* or *wrong* attribute, respectively.

Several tests can be performed on the collected corpus. It is possible to check in which contexts we find typing errors, which environment cause significant lags in typing time, etc. This paper concentrates on the modulation of inter-key duration, which we will call *lags.* The distribution of lags is not normal, leading us to use percentiles and medians rather than means and standard deviations. We first exclude all data from AMT workers that submitted more than five tasks, all keys that originate in *interrupted*, *corrected* or *wrong* words, all word-initial keys, and the top and bottom five percentiles of remaining lags. Following Gentner (1982) we build on the fact that the variance of keystrokes is reduced when preceding and following keys are taken into account as context. Like Weingarten et al. (2004) we use the preceding key for small corpora, but we also include the following key if at least ninety percent of every three-key sequences appear in the corpus at least five times. The median of each set of keys sequences is used as the *expected* lag of the the target key in that context. Our predicted value is the ratio between the actual lag and the expected lag, rather than the actual size of each lag. In this way, we can compare lag modulation across different words, and not limit ourselves to a specific key sequences. For example, if the lag for the letter 'e' in the context 'r*e*a' has an expected baseline of 220ms (based on all occurrences of the 'rea' in the corpus), but in a specific instance of the word 'great' we measure it to be 140ms long, we would like to explain why that particular 'e' is shorter, the value to predict being 140:220 (figure 1). Since the predicted value is the ratio, we can compare the ratios of different keystrokes, in different contexts.

| key | g | ⟷ | r | ⟷ | e | ⟷ | a | ⟷ | t |
|---|---|---|---|---|---|---|---|---|---|
| actual lag | | 100 | | 140 | | 30 | | 90 | |
| expected lag | | 210 | | 220 | | 100 | | 150 | |
| ratio | | 0.48 | | 0.64 | | 0.30 | | 0.61 | |

Figure 1: sample actual:expected ratios

## Study 1: Lexical frequency and neighborhood density in English

In the first study, we construct a typing time corpus for English, and use it to investigate the effects of neighborhood density and frequency on the typing-time. We predict a facilitatory effect of word frequency on its typing time. Additionally, we expect to find an effect of neighborhood density.

### Constructing an English typing-time corpus

The English typing-time corpus was built using AMT, using the procedure described above. Each AMT task was unique, but workers could participate in the study up to five times.

In order to choose the stimuli words to be typed, each word in the CMU Pronunciation Dictionary (Weide, 1998) was matched with its frequency and its most common letter case in the New York Times section of English Gigaword Third Edition (Graff, Kong, Chen, & Maeda, 2007): Gigaword-NYT. The corpus has two sections, which correspond to data collected using two different kinds of stimuli. Both tasks were used in order to calculate the expected lag of each keystroke in the context of one preceding and one following keystrokes.

In the first data collection task, AMT workers were requested to type in four randomly chosen words in each item. The words were independent from one another. Each word was in one of the top ten thousand lowercase words in Gigaword-NYT. A total of 475 AMT tasks were collected, and each took about two minutes to perform. No worker had to type the same word twice within the same hour.

The second data collection task required AMT workers to type in five words that form a coherent sentence, which was sampled from Gigaword-NYT. All sentences were exclusively in lowercase in the original corpus except for the first character, which was also changed into lowercase for the construction of the stimuli. The sentences were comprised only of words that are in the top five thousand most frequent words in Gigaword-NYT. No sentence had conjunctions or WH-words. Pronouns, if they appeared at all, occurred only before the verb. Each sentence had a verb and a noun following it. A total of 190 AMT tasks were collected, and each took about two minutes to perform. No worker had to type the same sentence twice within the same hour.

### Methods and materials

We investigate the effects of neighborhood density and frequency on the modulation of inter-key typing lag. A linear regression was used to estimate the predicted value, which was defined as the log ratio between a lag and its expected value. Only lags from the first section of the English typing-time corpus (words in isolation) were estimated. The key's position in the word, the predicted lag, AMT workers' typing rate across all items, their typing rate in the corresponding item and the logged predicted lag time were used as controls. The word frequencies used were the corresponding word counts in the NYT section of English Gigaword Third Edition (Graff et al., 2007): Gigaword-NYT. Two frequency measurement were tested. The first was the negative log unigram probability of that word: $-\log \Pr(word)$. The second word frequency measurement was based on the word lemmas: $-\log \Pr(lemma)$, calculated using WordNet (Miller, 1995).[2]

Neighborhood density was calculated using the CMU dictionary. We tested three variants of neighborhood density: the number of spelling neighbors (substitution of one letter), the number of phonological neighbors (substitution of one segment) and the number phonographic neighbors (substitution of one letter and one segment).

The linear regression model was selected using R's (R Development Core Team, 2010) `step()` function which uses AIC (Akaike, 1974) for model selection. The model was also re-evaluated using a mixed-effect model with worker and word as random effects. No significant changes to the significance and direction of the reported coefficients were found.

### Results and discussion

Both word and lemma frequency alone have a significant facilitatory effect on typing speed (words which are frequent or whose lemma is frequent are typed faster). However, in the final model only the frequency of lemma remained significant, as it masks the effect of the frequency of the word. The lemma unigram frequency has a significant ($p < 10^{-7}$) facilitatory effect and is significantly superior to word probability ($p < 0.02$).

All three neighborhood density measurements have a significant facilitatory effect on typing speed (words with a dense neighborhood are typed faster). However, in the

[2]If a word was ambiguous between two parts of speech, the shorter lemma was associated with the word.

final model only phonological density remains significant ($p < 0.001$). Phonological neighborhood density is not significantly better than spelling neighborhood density ($p = 0.097$) or phonographic neighborhood density ($p = 0.13$). The adjusted $R^2$ is 0.237

These results show that typing-time corpora are indeed sensitive to the well known effects of word frequency and neighborhood density. The fact that it is the frequency of lemmas rather than words suggests that lexicon access is active during typing, as shown in Weingarten et al. (2004). The fact that neighborhood density has a *facilitatory* effect is of particular importance, since it has been shown that in English a dense neighborhood facilitates productions whereas it inhibits comprehension (Vitevitch, 2002; Vitevitch & Luce, 1998). Therefore, although the typing task arguably involves both production and comprehension, the results suggest that this method is indeed tapping into the effects of production.

## Study 2: Lexical frequency in Hebrew

In the second study, we construct a typing time corpus for Hebrew. We use it to demonstrate that this paradigm is extensible to other languages, and can be collected outside AMT. We show that Hebrew demonstrates word frequency effects on typing-time.

### Constructing a Hebrew typing-time corpus

Hebrew orthography is different from that of English in several crucial aspects. It is written from right to left, it has no uppercase-lowercase distinction, and most importantly it does not incorporate most vowels. Furthermore, norms regarding the use of space are different — several very frequently occurring clitics (such as *ve* 'and') are glommed to the following word.

The Hebrew typing-time corpus was built using an online form, the results of which were collected by a web server.[3] One hundred unique tasks were generated, and each task was performed no more than three times, by different subjects. Subjects could participate in the study up to five times.

In order to choose the stimuli words to be typed, we collected 1300 articles from the *Haaretz*, a Hebrew news website. We calculated the frequency of each word and used Hspell (Har'El & Kenigsberg, 2006) to stem it from possible adjoining clitics. Word-frequencies were estimated using the same data from *Haaretz*. We calculated the expected lag of each keystroke in the context of one preceding keystroke.[4]

The data collection task was similar to the isolated word section of the English corpus described in study 1. Subjects were asked to type in five randomly chosen words in each item. The words were independent from one another. Each word was in one of the top five thousand in *Haaretz*. A total of 72 web tasks were collected, and each took about two minutes to perform. No worker had to type the same word twice.

[3]We did not use AMT because there are currently not enough native speakers of Hebrew in AMT

[4]There was not enough data to use the following key as well.

### Methods and materials

We investigate the effects of frequency on the modulation of inter-key typing lag. As in study 1, a linear regression was used to estimate the predicted value, which was defined as the log ratio between a lag and its expected value. Once again, the key's position in the word, the predicted lag, the subjects' typing rate across all items, their typing rate in the corresponding item and the logged predicted lag time were used as controls. We limited ourselves to words that had no clitics. The word frequencies used were the corresponding word counts *Haaretz*. Two frequency measurement were tested. The first was the negative log unigram probability of that word including its clitics when they occur, $-\log \Pr(clitics + word)$. The second word frequency measurement was based on the stemmed words (which still include morphological inflections, but not adjoining clitics) $-\log \Pr(word)$.

The linear regression model was evaluated as in Study 1. The linear regression model was selected using R's (R Development Core Team, 2010) `step()` function (Hastie & Pregibon, 1992) which uses AIC (Akaike, 1974) for model selection. The model was also re-evaluated using a mixed-effect model with worker and word as random effects. No significant changes to the significance and direction of the reported coefficients were found.

### Results and Discussion

Of the two word-frequency measurements, only the frequency of the word form that included its clitics came up significant $p < 0.05$. The frequency of the bare form did not come up significant even when we excluded the frequency of the cliticized form. The adjusted $R^2$ is 0.1462.

These results show that even with a much smaller typing-time corpus, frequency effects can be seen.

## General Discussion

The experimental results shown in both studies provide strong support of our proposal that typing time corpora can provide a simple method to investigate linguistic performance. Further investigation is required to assess the many different ways in which production is similar or different across the typed and spoken modalities.

Study 1 shows that the reduction of frequent words, an effect shown by both laboratory experiments and phonetically tagged corpora, has a corollary in typing time which can be replicated using our corpus. It also shows that a facilitatory effect of neighborhood density can be observed using our corpus, which shows that it patterns with production rather than comprehension.

Study 2 demonstrates that this methodology can be easily extended to other, less studied languages. The results show a shortening of typing lag in more frequent words in Hebrew, as was shown for English in Study 1. This demonstrates that the method is applicable to new languages, even those with non-Roman orthographies.

## Acknowledgments

This research was partially supported by the NSF via award IIS-0624345. Special thanks to Dan Jurafsky, Roey Gafter, Chigusa Kurumada, Victor Kuperman, Matthew Adams and Meghan Sumner.

## References

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review*, *14*(3), 455–459.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*, 3048–3058.

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

Callison-Burch, C. (2009). Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of EMNLP 2009.*

Colowick, S. M., & Pool, J. (2007). Disambiguating for the web: a test of two methods. In *Proceedings of K-CAP '07* (pp. 173–174). ACM.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing *the* as *thee* to signal problems in speaking. *Cognition*, *62*, 151–167.

Gentner, D. R. (1982). Evidence against a central control model of timing in typing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(6), 793–810.

Gentner, D. R., Larochelle, S., & Grudin, J. (1988). Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology*, *20*(4), 524–548.

Godfrey, J. J., & Holliman, E. (1997). *Switchboard-1, Release 2.* Linguistic Data Consortium, Philadelphia.

Graff, D., Kong, J., Chen, K., & Maeda, K. (2007). *English gigaword third edition.* Linguistic Data Consortium, Philadelphia.

Har'El, N., & Kenigsberg, D. (2006). *Hspell.* `http://hspell.ivrix.org.il/`.

Hastie, T. J., & Pregibon, D. (1992). Generalized linear models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (chap. 6). Pacific Grove, CA: Wadsworth and Brooks / Cole.

Kohler, K., Pätzold, M., & Simpson, A. (1995). *From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech. AIPUK 29.* Kiel: IPDS.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Commun. ACM*, *38*(11), 39–41.

Oostdijk, N. (2000). The spoken dutch corpus project. *ELRA newsletter*(5), 4–8.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, *37*(3), 382–410.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye corpus of conversational speech, 2nd release.* Department of Psychology, Ohio State University.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Available from `http://www.R-project.org`

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of EMNLP 2008* (pp. 254–263).

Van Son, R. J. J. H., & Van Santen, J. P. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*, 100–123.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325.

Weide, R. (1998). *The CMU pronunciation dictionary, release 0.6.* (Carnegie Mellon University)

Weingarten, R., Nottbusch, G., & Will, U. (2004). Morphemes, syllables and graphemes in written word production. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (Vol. 157, pp. 529–572). Berlin: Mouton de Gruyter.

Zesiger, P., Orliaguet, J., Boë, L., & Mounoud, P. (1994). The influence of syllabic structure in handwriting and typing production. In C. Faure, P. Keuss, G. Lorette, & A. Vinter (Eds.), *Advances in handwriting and drawing: a multidisplinary approach* (pp. 389–401). Paris: Europia.