Word competition: an entropy-based approach in the DIANA model of human word comprehension

Louis ten Bosch[1], *Lou Boves*[1]

[1]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

l.tenBosch@let.ru.nl, l.boves@ru.nl

Abstract

We discuss the role of entropy of the set of unfolding word candidates in the context of DIANA, a computational model of human auditory speech comprehension. DIANA consists of three major interacting components: Activation, Decision and Execution. The Activation component computes activations of word candidates that change over time as a function of the unfolding audio input. The resulting set of word candidate activations can be associated with an entropy that is related to difficulty of the decision when one of these candidates must be selected at time T. The paper presents the close relation between entropy measures and the between-word competition during the unfolding of the auditory stimuli, and at the end of the stimuli if no decision could be made before stimulus offset. We present a way for computing the entropy that takes into account linguistic-phonetic constraints that play a role in speech comprehension and in lexical decision experiments. Using the BALDEY data set and linear mixed effects regression models for RT, we show that entropy measures explain differences between RTs of words with different morphological structure.

Index Terms: entropy, reaction times, psycholinguistics, cognitive processes, auditory lexical decision

1. Introduction

How listeners comprehend speech is only partly understood. Influential theories and models of human speech comprehension (e.g., TRACE [1], Shortlist-B [2], SpeM [3]) assume that during the unfolding of the speech signal a competition takes place between words on the basis of their activation, which is in turn determined by the degree of match between the input speech signal and some internal representation of the words. In most models, the focus is on a particular part of the processes that take place between the start of a speech signal and the eventual overt decision related to the status of a stimulus, such as a button press in a lexical decision experiment or eye fixations in a visual world paradigm.

Importantly, in most of the classical computational models of speech comprehension, the processing of the actual audio signal is mostly sidestepped, either by assuming a (handcrafted, idealized) spectral vector representation (such as done in TRACE), taking a handcrafted symbolic sequence (Shortlist [4]), or input confusion probability tables (Shortlist-B) derived from human phone classification errors in a large-scale gating experiment based on diphones [5]. This sidestepping is often defended on the basis of our lack of knowledge about how humans process input audio signals in terms of activating prelexical representations (e.g., [2]). Due to the recent advances in cognitive neuroscience and speech decoding techniques, however. that argument is now losing its appeal. For this new line of research in this field see, e.g., [6].

Recent models such as FineTracker [7], DIANA [8, 9], LDL [6] and EARSHOT [10] aim to model human speech processing by taking the actual waveform as input. LDL, DIANA, and EARSHOT also avoid the *necessity* of assuming a segmental 'prelexical' layer between the audio signal and the lexical layer. In a prelexical layer, smaller subword units (e.g. phoneme-sized) are postulated to be activated based on their match with the audio signal, after which the lexicon is accessed based on the activations in the prelexical layer. The exact representation of the audio, the exact assumptions about the status of the prelexical layer and the interaction between the layers in computational models of human speech processing are topics of ongoing debates ([11, 12]). LDL, EARSHOT and DIANA show that a discrete symbolic prelexical representation (that causes problems in FineTracker) is not necessary. DIANA differs from LDL in that it is a process-oriented model, rather than a model based on constructing linear mappings between vector spaces [13]. DIANA also differs from EARSHOT: DIANA models human word comprehension in terms of activation, decision and execution, while EARSHOT models the mapping between audio input and distributed semantic representations using a interactive activation-based neural network [10].

In this paper, we focus on the DIANA model. The primary motivation is that DIANA is very much oriented to modelling underlying processes (in particular Activation and Decision processes, see below). Using DIANA, we focus on the activations of word hypotheses as a function of time and the relation with human reaction times in lexical decision experiments, with the aim to further improve the Decision component in DIANA.

DIANA (e.g., [14]) has been used to simulate RTs in the Dutch BALDEY database [15] and the large-scale North American English data (MALD) [16], for compounds [17], to differentiate predictors in regression models [18], and for modelling the cross-entropy as a significant predictor for modeling ERP components during comprehension of continuous speech in a recent EEG experiment [19].

DIANA consists of three interacting components: Activation, Decision and Execution [8]. The Activation component takes into account knowledge about phone activations in the auditory cortex [20, 21] by mapping the input audio signal to sequences of Spectro-Temporal Receptive Fields (STRFs). This component computes activations of word candidates that change over time as the input $audio[0:T]$ unfolds as a function of T. The resulting set of word hypotheses is input for DIANA's Decision component. In the current Decision component, a word candidate is declared a 'winning' word as soon as the activation of the top hypothesis exceeds the activation of the runner-up by a threshold ($\theta > 0$). The value of θ determines the speed-accuracy trade-off in DIANA's recognition results [18]. The set of competing word candidates can be associated with an entropy measure that predicts the time it takes to decide which candidate to select (after applying the Luce rule: $p_i = \exp(-c \cdot act_i)/\sum_j \exp(-c \cdot act_j)$). As a result, there is a close relation between (a) the entropy in the word search during the stimulus unfolding, (b) the likelihood of any hypothesis to be declared a winning hypothesis (the smaller the entropy, the smaller the number of candidates to choose from and the faster the choice), (c) DIANA's way to compute these activations from the speech signal in the Activation component, and (d) the role of DIANA's entropy for modelling human reaction times in lexical decision experiments.

In this paper, we will illustrate this relation by showing (a) the relation between DIANA and this entropy, and (b) that this entropy is a significant predictor in linear mixed regression models simulating observed reaction times in lexical decision experiments. That there is such a relation is suggested by other research. [19]'s ERP study using over 2 million EEG traces has shown that DIANA's activation of word cohorts for gated stimuli $audio[0 : T]$ for varying T is closely related to the cross entropy between two probability density functions pdf_0 and pdf_T. Here pdf_0 denotes the pdf on the lexicon entirely based on the expectation from the word's precontext, while pdf_T denotes the pdf based on DIANA's bottom-up activations based on the audio stretch $audio[0 : T]$.

By investigating the role of entropy during the audio presentation, we provide a basis for further improvement of the Decision component in DIANA, and advancing DIANA more towards a process model.

2. Output of DIANA's Activation component

The Activation component in DIANA used in previous experiments (e.g., [8, 16]) used software developed in the context of Automatic Speech Recognition (HTK toolkit [22]) for the implementation of the Activation component. Recently, a new implementation of this component has been completed, using software from the Kaldi toolkit [23]. The most important advantage of Kaldi over HTK is the availability of more flexible tools for handling the lattices that contain the dynamically changing activations of word and pseudoword candidates as the acoustic stimulus unfolds. Another advantage of Kaldi is that it allows using lexicons with an essentially unlimited number of entries, whereas the lexicon size in the HTK-based implementation was limited to about 25,000 entries. The experiments reported here use a lexicon with over 500,000 word forms, including derivations and inflexions, compounds, names, etc.

DIANA and similar models play a large role in the understanding what human listeners exactly do during speech perception. Despite the fact that lexical decision experiments have been used time and again to investigate issues related to spoken word comprehension, it is not well known what strategies participants use to perform the task. One question is whether they aim to identify one specific word (if the stimulus is a word), namely the stimulus word, or whether they decide for 'word' when any word they know matches sufficiently closely with the acoustic stimulus. In any case, a plausible Activation component cannot be limited to estimating the match for the unique word intended by the experimenter. Since participants do not know that word beforehand, they will activate all phonetically similar words. If it is the goal to recognize a unique word, phonetically similar words will act as competitors (as they might do in a daily conversation) that 'steal' a part of the total activation from the intended word. However, if the goal is to decide whether the stimulus is any word, the activations of the competitors actually add to the 'wordiness' of the stimulus.

During the unfolding of the acoustic stimuli the Kaldi lattices may contain multiple hypotheses that actually refer to the same entry in the (mental) lexicon. In calculating the measures that best describe the competition, all 'redundant' hypotheses are dropped; of the hypotheses that point to the same lexicon entry only the one with the highest activation score is kept. Still, in almost all stimuli several hypotheses pointing to 'competing' words remain, possibly along with hypotheses that point to a pseudoword. With these data we compute three different scores: (a) the formal entropy of the remaining hypotheses, (b) the total 'wordiness' of the stimulus, defined as the sum of the activations of all word hypotheses, and (c) the distance between the activation of the winner and the first hypothesis that is of a different type (word or pseudoword) than the winner.

As a result of this decoding, for each moment T between word onset and word offset, a list of hypotheses (word cohorts) is available in combination with their bottom-up scores, the entropy, the wordiness, and the score difference.

Figures 1, 2, 3, 4 show DIANA's online activation over time for two words (poverheid and aalscholver) and two nonwords (aafpreel and zwokkelijk). The horizontal axis displays time (in ms). The vertical axis displays the acoustic score. For the first three figures, the winning hypothesis accrues acoustic support from the audio signal which leads to the increasing line. Competing words deviate at specific time points from the winning hypothesis. These plots show that the concurrence of words is not evenly spread over time. The branches that lead to runner up hypotheses are the subsymbolic audio-signal-based equivalents of an increasing symbolic phone string that grows from left to right. These plots therefore provide a realistic insight in the actual bottom-up competition between words, given a realistically sized (over 300k) lexicon.

The vertical axis in fig. 4 is chosen in a different way: in this figure, the axis shows the acoustic bottom-up score but averaged over the number of acoustic 10ms portions seen so far. On top the confidence interval is shown, indicating when the winning hypothesis is significantly different in score compared to all other word hypotheses.

These figures clearly show that the acoustically supported hypotheses branch off from the winning branch, or they initiate halfway with a lower activation (such as in fig. 4). From the acoustical point of view, it is not at all comparable to a rat race with dropouts. Instead, it is a dynamic come-and-go.

3. Entropy

In order to compute the entropy during the word decoding phase, we examined the competition of all possible word starts that match the gated signal $audio[0 : T]$ with T ranging from 110 ms to word offset. For each T we obtain a ranked list of partial hypotheses, each either belonging to a possible word start or the start of a non-word (pseudo word). In this list, different types of word 'neighborhoods' can be defined based on the form-relation between hypotheses: any two hypotheses can be 'nested' (i.e., one is a substring of the other, e.g., 'rUdi' and 'rUdim'), or non-nested. Within each list, each variant that was nested within any variant with a higher activation was removed. Next, the entropy was computed from the activations a_i, after applying the Luce rule $p_i = \exp(ca_i)/\sum_j \exp(c \cdot a_j)$, via $H = -\sum p_i \log(p_i)$. This procedure was repeated for each time stamp, for all 5541 BALDEY stimuli. The value of $c > 0$ is estimated to be 0.1, by using the word-confidence estimation approach described in [24].

In a similar vein, two other parameters are computed for

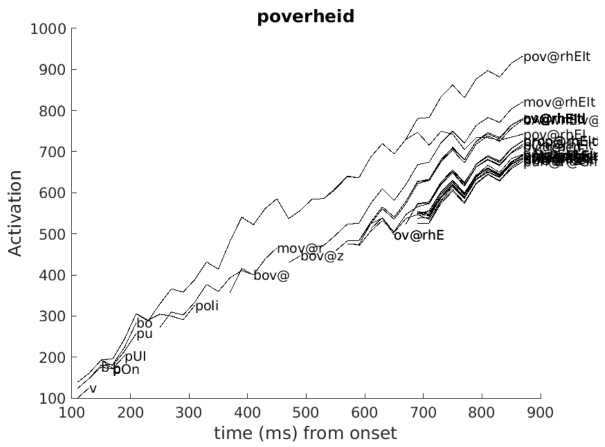

Figure 1: *Activation as a function of time for the word poverheid.*

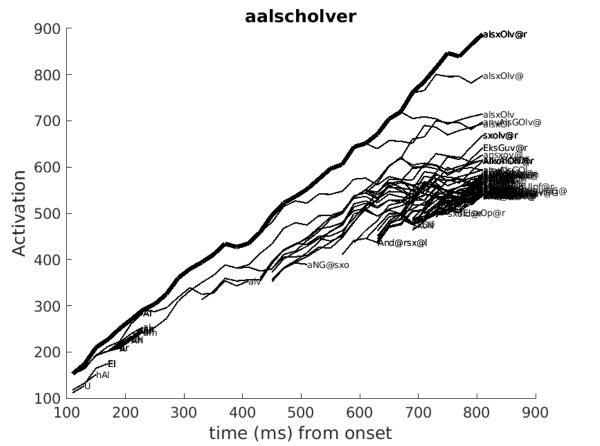

Figure 2: *Activation as a function of time for the word aalscholver.*

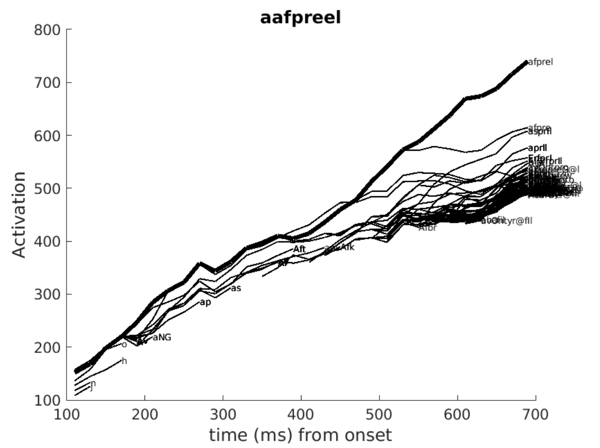

Figure 3: *Activation as a function of time for the pseudo word aafpreel.*

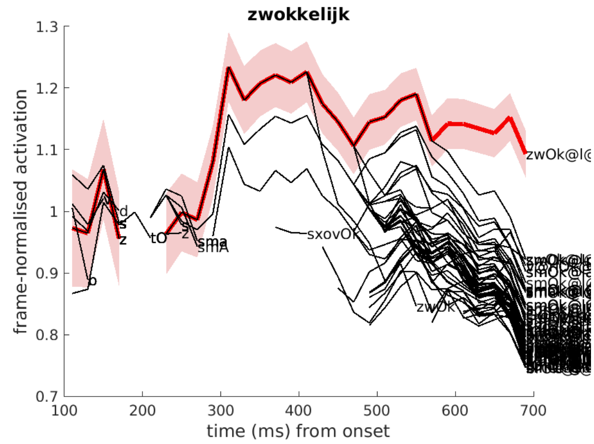

Figure 4: *Activation as a function of time for the pseudo word zwokkelijk. For the eventually winning hypothesis, the confidence interval is added.*

each T until stimulus offset: the proportion of the total probability assigned to word hypotheses (wordActivation) and the difference between the activations of the best word-based hypothesis and the best non-word-based hypothesis (typeDifference). All these parameters change as a function of T between 0 and stimulus offset.

4. Regression analysis

4.1. Removing local speed effects from the participant RTs

In experiments with speeded RTs, the RTs are the result of short-term stimulus dependent effects and longer term effects such as participant-dependent effects and mid-term behavioral effects (the 'local speed effect'). In order to be able to remove local speed effects in the raw RT sequences produced by the participants, needed for obtaining reliable estimates of the average RT of all stimuli, we preprocessed each of the $20 \cdot 10$ BALDEY RT sequences individually, in the same way as is done in [25]. This step mainly deals with missing RTs.

4.2. Set-up

In order to investigate the effect of entropy, wordActivation and typeDistance we analysed the log-transformed reaction times (logRT) by using linear mixed effects models using these predictors. As a control predictor we also included maRT, which is a weighted version of the conventional 'previous RT' (see [18, 25] for more details). In addition, we included the predictor prevBVis, which is based on the Visibility Graph Analysis (VGA) [26] applied on the logRT sequence. These two predictors model the 'local speed', i.e., the local trend that determines reaction time sequences to a considerable extent, and are usually highly significant in many RT-based regression models ([18]).

Other main predictors are logwdur (log-transformed word duration), logFreq, session, trial, wordclass, and compound-type. The predictor wordclass is a factor with three levels (adjective, noun, and verb, with 22154, 48861 and 32917 tokens, respectively). 'Adjectives' are on the intercept. Compound-type distinguishes four types of compounds (simple, adj+noun, noun+adj, and noun+noun). Here, 'simple' is on the intercept. Since the correlation between logwdur and entropy was approx. 0.5, entropy was residualized over log word duration, with entrlogwdur as new predictor.

4.3. Regression results

Not all models with complex random structure converged. In the final model, the two-way interaction between logwdur and log frequency is kept, and maRT is put as random slope under subject without correlation with the intercept. Table 1 presents the results of model m2a, the best that converged using the guidelines in [27, 28]). The AIC of the model shown equals -753.2811.

```
m2a = lmer(logRT ~ logwdur*mylogFreq
+entrlowdur+wordActivation+session+trial
+wordclass+maRT+prevBVis+compound_type
+(1|word)+(1 |subject)+(0+maRT2 |subject),
data=data4[data4$response == "correct",])
```

RT_{onset}			
	Estimate	Std. Error	t value
(Intercept)	6.742e-01	3.265e-01	2.065
logwdur	3.306e-01	6.513e-03	50.760
logFreq	7.011e-02	1.135e-02	6.178
entrlogwdur	2.354e-02	2.039e-03	11.548
wordActivation	5.559e-02	9.084e-03	6.119
session	2.564e-03	2.919e-04	8.783
trial	2.672e-05	4.883e-06	5.472
wordclass(nom)	6.933e-03	2.965e-03	2.338
wordclass(verb)	3.106e-02	2.951e-03	10.526
maRT	6.000e-01	4.299e-02	13.958
prevBVis	2.651e-03	2.883e-04	9.196
compoundtypeA+N	-3.023e-02	8.855e-03	-3.414
compoundtypeN+A	-7.375e-03	1.018e-02	-0.724
compoundtypeN+N	4.197e-03	4.236e-03	0.991
logwdur:logFreq	-1.277e-02	1.785e-03	-7.153

Table 1: *Output of regression model modelling RT using entrlogwdur, which is the entropy residualized over log word duration, and wordActivation.*

4.4. Findings

Table 1 provides β's and t-values; under the assumption of the t-distribution, all values of $|t| > 1.96$ are considered to flag significance. As always, word duration is one of the most significant predictors of reaction time. The β of log word freq (logFreq) is positive but this seemingly delaying effect is compensated by the significant negative interaction between log word duration and log frequency. Later sessions and trials lead to slower RTs. Compared to adjectives, nouns and verbs slow down. Noun-noun combinations produce the slowest RTs among compounds (not surprising given their processing ambiguity). The 'local trend'-related control predictors maRT and prevBVis are often highly significant with $\beta > 0$, again showing the usual substantial local speed effect [25].

The effects of entropy and of wordActivation are the most interesting for the interpretation of competition-based measures for the RT and for the further improvement of the Decision component in DIANA. It appears that the entropy (when residualised over word duration) slows down the reaction time. This finding supports the theoretical underpinning used in DIANA to explain RT in auditory lexical decision experiments as a sum of three contributions: (a) the word duration as a result of the unfolding of the auditory stimulus, (b) the choice-RT, i.e., the time it takes to make a choice from the set of hypotheses eligible at stimulus offset, and (c) the time it takes to execute a command such as a button press. In [18], [17] is it stipulated that the choice-RT equals a certain constant times the entropy at word offset, in line with Hick's law [29]. The model presented in table 1 shows this effect. The β in the table differs from earlier estimations [8], likely because the closely related predictor wordActivation (derived from the same process) is weakly correlated with word duration (0.31) and therefore affects the precise effect of entropy as main predictor. The positive β for word-Activation shows that the more probability mass is attributed to word hypotheses, the slower the decision will be. This seems counter-intuitive, since one may expect that the larger the probability mass taken by one option, the faster the decision should be. However, since wordActivation does not specifically take into account whether the winning hypothesis is included in this set, it may well be that the larger word 'mass' delays decision, since the majority in DIANA's decoding lexicon consists of real words, thereby hampering (biasing) the decision for a non-word judgement. If so, this would indicate that the precise mechanism regulating the prior probability of a lexical or non-lexical decoding result in DIANA must be improved.

5. Discussion and conclusion

Several recent computational models of human word comprehension take audio into account, but as far as we know DIANA is the only model (see, e.g., [16]) that takes in the actual speech signal and that implements an Activation, Decision, and Execution component. For several years, the interaction between the Activation and Decision components has been a focus for further improvement. The time-varying entropy changes during the stimulus presentation (highest at stimulus onset, decreasing over time until stimulus offset). Its eventual value at stimulus offset may be zero, in which case DIANA does not add a choiceRT to the overall RT, but often the entropy is positive, in which case DIANA adds a non-zero choiceRT. This mechanism, inspired by Hick's law, is well known in the literature on human decision making. In this paper we present results showing that entropy, next to the conventional control predictors, is a useful measure in two ways: (a) as predictor in a data-oriented lmer() model, and (b) as theoretical underpinning of DIANA's choiceRT rule, and thereby closely related to the way DIANA processes the speech signal and produces dynamically changing lists of word (cohort) hypotheses. There is a close connection with the cross entropy between the two probability density functions pdf_0 and pdf_T which unfold over time ([19]).

The fact that the predictor wordActivation has a significantly positive β in the regression model 1 suggests that the entropy measure applied here can still be improved. The current measure is based on the identification of individual word hypotheses, but the participant's task in lexical decision is not about which word it is, but whether it *is* a word or not. That means that the true competition is not between individual hypotheses, but between groups of hypotheses. Our plan for the near future is to extend this research to the North-American lexical decision database MALD [30].

6. Acknowledgements

The development of DIANA was funded by an NWO VICI grant awarded to Mirjam Ernestus.

7. References

[1] J. L. McClelland and J. L. Elman, "The trace model of speech perception," *Cognitive psychology*, vol. 18, no. 1, pp. 1–86, 1986.

[2] D. Norris and J. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological Review*, vol. 115, pp. 357 – 395, 2008.

[3] O. Scharenborg, D. Norris, L. ten Bosch, and J. McQueen, "How should a speech recognizer work?" *Cognitive Science*, vol. 29, pp. 867 – 918, 2005.

[4] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition*, vol. 52, pp. 189–234, 1994.

[5] R. Smits, N. Warner, J. McQueen, and A. Cutler, "Unfolding of phonetic information over time: A database of Dutch diphone perception," *Journal of the Acoustical Society of America*, vol. 113, pp. 563–574, 2003.

[6] D. Arnold, F. Tomaschek, K. Sering, F. Lopez, and R. Baayen, "Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit," *PLoS ONE*, vol. 12, no. 4, pp. 1 – 16, 2017.

[7] O. Scharenborg, "Modeling the use of durational information in human spoken-word recognition," *Journal of the Acoustical Society of America*, vol. 127, pp. 3758 – 3770, 2010.

[8] L. ten Bosch, L. Boves, and M. Ernestus, "Comparing reaction time sequences from human participants and computational models," in *Proceedings of Interspeech*, Singapore, 2014, pp. 462 – 466.

[9] L. ten Bosch, M. Ernestus, and L. Boves, "Analyzing reaction time sequences from human participants in auditory experiments," in *Proc. Interspeech 2018*, 2018, pp. 971–975.

[10] J. S. Magnuson, H. You, S. Luthra, M. Li, H. Nam, M. Escabí, K. Brown, P. D. Allopenna, R. M. Theodore, N. Monto, and J. G. Rueckl, "EARSHOT: A minimal neural network model of incremental human speech recognition," *Cognitive Science*, vol. 44, no. 4, p. e12823, 2020.

[11] D. Norris, J. M. McQueen, and A. Cutler, "Commentary on "interaction in spoken word recognition models"," *Frontiers in Psychology*, vol. 9, 2018.

[12] J. S. Magnuson, D. Mirman, S. Luthra, T. Strauss, and H. D. Harris, "Interaction in spoken word recognition models: Feedback helps," *Frontiers in Psychology*, vol. 9, 2018.

[13] A. Lugmayr, B. Stockleben, C. Scheib, and M. Mailaparampil, "Cognitive big data - survey and review on big data research and it's implications: What is really new in big data?" *Journal of Knowledge Management*, vol. 21, 02 2017.

[14] L. ten Bosch, L. Boves, and M. Ernestus, "Towards an end-to-end computational model of speech comprehension: Simulating a lexical decision task," in *Proceedings of Interspeech*, Lyon, France, 2013, pp. 0000–0000.

[15] M. Ernestus and A. Cutler, "BALDEY: A database of auditory lexical decisions," *Quarterly Journal of Experimental Psychology*, vol. Advance online publication, 2015.

[16] F. Nenadić, L. ten Bosch, and B. V. Tucker, "Implementing DIANA to model isolated auditory word recognition in english," in *Proc. Interspeech 2018*, 2018, pp. 3772–3776.

[17] L. ten Bosch, L. Boves, and M. Ernestus, "The recognition of compounds: a computational account," in *Proceedings of Interspeech, Sept. 20–24*, Stockholm, 2017, pp. 000 – 000.

[18] ——, "Combining data-oriented and process-oriented approaches to modeling reaction time data," in *Proceedings of Interspeech, Sept.*, San Francisco, 2016, pp. 000 – 000.

[19] M. Bentum, L. ten Bosch, A. van den Bosch, and M. Ernestus, "Listening with Great Expectations: An Investigation of Word Form Anticipations in Naturalistic Speech," in *Proc. Interspeech 2019*, 2019, pp. 2265–2269. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2741

[20] N. Mesgarani, C. Cheung, K. Johnson, and E. F. Chang, "Phonetic feature encoding in human superior temporal gyrus," *Science*, vol. 343, no. 6174, pp. 1006–1010, 2014.

[21] H. Akbari, B. Khalighinejad, J. L. Herrero, A. D. Mehta, and N. Mesgarani, "Towards reconstructing intelligible speech from the human auditory cortex," *Nature, Scientific Reports*, vol. 9, no. 874, pp. 1 – 10, 2019. [Online]. Available: https://doi.org/10.1038/s41598-018-37359-z

[22] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, "The HTK book (for HTK version 3.4)," Cambridge University Engineering Department, Cambridge, UK, Tech. Rep., 2009.

[23] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, 2011, iEEE Catalog No.: CFP11SRW-USB.

[24] F. Wessel, R. Schlüter, K. Macherey, and H. Ney, "Confidence measures for large vocabulary continuous speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 288 – 298, 2001.

[25] L. ten Bosch, L. Boves, and K. Mulder, "Analyzing reaction time and error sequences in lexical decision experiments," in *Proceedings Interspeech*, Graz, Austria, 2019, pp. 2280 – 2284.

[26] A. Turner, M. Doxa, D. O'Sullivan, and A. Penn, "From isovists to visibility graphs: a methodology for the analysis of architectural space," *Environment and Planning B*, vol. 28, no. 1, pp. 103–121, 2001.

[27] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, "Parsimonious mixed models," pp. 1 – 26, 2018. [Online]. Available: https://arxiv.org/pdf/1506.04967.pdf

[28] L. Meteyard and R. A. Davies, "Best practice guidance for linear mixed-effects models in psychological science," *JML*, vol. 112, 2020.

[29] R. W. Proctor and D. W. Schneider, "Hick's law for choice reaction time: A review," *Quarterly Journal of Experimental Psychology*, vol. 71, no. 6, pp. 1281–1299, 2018.

[30] L. ten Bosch, L. Boves, B. Tucker, and M. Ernestus, "DIANA: towards computational modeling reaction times in lexical decision in north american english," in *Proceedings of Interspeech*, Dresden, Germany, 2015.