Title
Predictability effects in adult-directed and infant-directed speech: Does the listener matter?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Pate, John
Goldwater, Sharon

Publication Date
2011

Peer reviewed

Predictability effects in adult-directed and infant-directed speech:
Does the listener matter?

John K. Pate (j.k.pate@sms.ed.ac.uk) and Sharon Goldwater (sgwater@inf.ed.ac.uk)
School of Informatics, University of Edinburgh, 10 Crichton St., Edinburgh EH8 9AB, UK

Abstract

A well-known effect in speech production is that more predictable words tend to be phonetically reduced. Recent work has suggested that predictability effects result from hardwired properties of the language production system, rather than active modulation by the talker to accommodate the listener. However, these studies investigated only minor manipulations of listener characteristics. Here, we examine predictability effects with two very different listener populations: adults and preverbal infants. Using mixed effects regressions on spontaneous speech corpora, we compare the effect of word frequency, probability in context, and previous mention on word duration in adult-directed and infant-directed speech. We find that the effects of preceding context and word frequency differ according to listener. Contrary to previous work, these results suggest that talkers do modulate the phonetic effects of predictability based on listener characteristics. To our knowledge, this study is also the first published analysis of predictability effects in infant-directed speech.

Introduction

It has long been known that more predictable words (e.g., those with higher frequency or previous mentions in the discourse) tend to be phonetically reduced (e.g., with shorter duration and less distinct vowels). Recent investigations into this phenomenon have focused on an information-theoretic explanation known as the Smooth Signal hypothesis (Aylett & Turk, 2004) or the Uniform Information Density hypothesis (Frank & Jaeger, 2008). Under this theory, speech is viewed as a method of communicating information across a noisy channel. The most efficient encoding of information is one in which there is a constant cost (in terms of time or effort) for transmitting a single unit of information, so when words are more predictable (conveying less information), a rational speaker will tend to pronounce them more quickly (reducing time) or articulate less clearly (reducing effort).

This theory assumes that speech production is optimized for efficient communication, but does not specify the source of this optimization. One possibility is that optimization is strictly offline: the talker's speech production system is optimized for communication across many talker-listener pairs through talker-based mechanisms such as priming, lexical access, and articulatory practice. Another possibility is that some or even most optimization occurs online, with talkers adapting their pronunciation to the immediate needs of their listeners or communicative goals, for example by reducing the pronunciation of a word only if they believe that the current listener would find it predictable.

Earlier work on predictability effects often assumed online modulation of these effects in response to listeners' needs (e.g., Lindblom, 1990), but more recent studies have cast doubt on this hypothesis. For example, Bard et al. (2000) analyzed data from dialogues between pairs of subjects negotiating a route on a map, where the subjects' maps differed slightly. Bard et al. found that talkers reduced the intelligibility and duration of words they had previously mentioned, regardless of whether the current listener had heard or seen these words before. Bard et al. also found that talkers reduced words that had previously been mentioned only by the listener, but they suggest that this effect could be the result of talkers forgetting who mentioned the word previously, rather than explicitly adjusting to listener knowledge. Put another way, talker-based priming mechanisms, rather than listener modeling, are sufficient to explain the result.

While Bard et al.'s study suggests that talkers do not maintain detailed models of listeners' discourse knowledge to determine word predictability (and thus pronunciation), it is still possible that talkers modulate predictability effects for the benefit of the listener in response to more general knowledge about listener characteristics, such as the listener's overall linguistic competence, or communicative goals.

In this paper, we present evidence in favor of this weaker version of listener-based pronunciation adaptation. We do so by examining the effects of predictability on word duration in a type of speech that, to our knowledge, has not been studied previously for this purpose: infant-directed speech (IDS). By comparing predictability effects in IDS with those in adult-directed speech (ADS), we aim to tease apart two hypotheses regarding the role of the listener in determining word predictability and associated reductions in pronunciation:

- **Hypothesis 1:** All effects of predictability on pronunciation are hard-wired into the language production system (or are otherwise talker-based), and so are not subject to modulation in response to the listener's needs.
- **Hypothesis 2:** Some predictability effects are subject to modulation by the talker according to general knowledge about the listener's needs and communicative goals.

If Hypothesis 1 is correct, we would expect no differences between ADS and IDS in terms of which predictability factors affect word duration, or the direction of the effects. Talkers may speak more slowly in IDS, or stretch all the predictability effects in the same direction because communication with infants represents a noisier channel. But even if IDS talkers are not interested in communicating efficiently with a preverbal infant, or assume little or no language knowledge on the part of the infant, predictability effects persist because they are results of properties of the talker alone.

Under Hypothesis 2, however, we would expect some differences in predictability effects between ADS and IDS. For example, since the listener is a pre-verbal infant, talkers may assume that the listener has a poor model of the probability of words in context, and so no longer reduce (or overarticulate) words according to their probability in context. Or, if efficient communication of a message is not a goal in IDS, then we would expect to see only predictability effects that the talker has no control over, such as those due to lexical access processes.

Our study is inspired by that of Bell, Brenier, Gregory, Girand, and Jurafsky (2009), who compared the effects on function and content word duration of several predictability factors, including word frequency, first mention, and probability in context. However, our study differs from Bell et al. (2009) in several respects: first, rather than comparing function and content words, we compare ADS and IDS. Second, whereas Bell et al. (2009) used a linear regression model to control for confounds such as talker age, speech rate, and syllable count, we use a more sophisticated mixed effects regression, which controls as well for correlations between items due to talker and sentence.

We present three regression studies. First, we perform separate regressions on the ADS and IDS data to determine which predictability factors significantly affect word duration in each case. In ADS we find significant independent shortening effects of higher overall frequency, higher conditional probability based on either preceding or following context, and previous mention in the discourse. In IDS, we find all of these effects except the effect of preceding context. A third regression with pooled ADS and IDS data and an additional factor for Speech Type (ADS or IDS) confirms that the effect of preceding context is at least significantly weaker in IDS, if not absent altogether. It also reveals that the shortening effect of word frequency is significantly stronger in IDS. Additionally, while, in common with Bard et al. (2000), we find no modulation according to listener of First Mention alone, we do find a modulation according to listener in an interaction between First Mention and probability in context.

As ADS and IDS share the same population of talkers but do not share the same population of listeners, these results suggest that at least some of the mechanisms behind predictability effects are listener-based. Moreover, as detailed in the Discussion Section, the particular differences found between ADS and IDS suggest that those effects which disappear in IDS are precisely those effects which we would expect to be primarily listener-based.

Materials and Methods

Data

We use data from two corpora in our study: `swbdnxt` (Calhoun et al., 2010), an edition of the Switchboard corpus of telephone conversations between adults, and `Large Brent` (Rytting, Brew, & Fosler-Lussier, 2010), a subset of the Brent corpus (Brent & Siskind, 2001) of spontanous infant-directed speech. We describe these corpora and the data extracted from them below.

Table 1: Statistics for the two datasets used on this paper.

	# Sent	# Words	$\frac{\text{Word}}{\text{Sent.}}$	# Talkers
ADS	2,273	16,301	7.2	73
IDS	2,254	14,148	6.3	4

swbdnxt `swbdnxt` is an edition of the Switchboard corpus of telephone dialogues between adults. It integrates several levels of annotation produced by different groups since the original Switchboard release. These include prosodic (ToBI) and syntactic annotations, as well as a phonetic alignment created by correcting the output of a forced alignment produced using a pronunciation dictionary.

To create the dataset for our experiments, we began with the 75 conversations that are annotated with ToBI, Mississippi state phonetic alignments, and Penn Treebank POS tags and parses. We discarded one conversation due to inconsistent annotation, and one talker due to missing metadata. The resulting corpus contained $12,140$ sentences ($99,965$ words), from which we removed all sentences longer than 20 words and shorter than 3 words, eliminating $4,704$ sentences ($39,452$ words). We split the remaining sentences into 80% training, 10% development, and 10% test sets in anticipation of future work (not described here). For this study, we use only the training set, and only talkers of side A (to avoid unhandled correlations between talkers in the same conversation), a total of 23,638 words from 2,608 sentences. Following Bell et al. (2009), we discard all words adjacent to disfluencies (identified by the POS tags "UH" and "XX") to avoid the complicated effects of disfluencies on word duration. Also following Bell et al. (2009), who note that short prosodic phrases are typically formulaic discourse responses (i.e. "oh good grief"), we discard prosodic phrases that are three words or shorter. Table 1 shows statistics for the final `swbdnxt` corpus.

Large Brent `Large Brent` is a subset of the Brent Corpus of spontaneous IDS collected in a naturalistic setting. It consists of the mothers' utterances from four mother-infant dyads, and has a forced phone alignment based on a modified version of the CMU pronunciation dictionary. Details of the corpus and alignment can be found in Rytting et al. (2010). `Large Brent` has a 90%/10% train/test partition; for this study we use only the training partition, which contains $22,226$ words from $7,030$ sentences. Rytting et al. (2010) have already excluded utterances containing partial or unintelligible words, so we made no further effort to handle disfluencies.

Unlike `swbdnxt`, this corpus does not include talker's ages; since all talkers are new mothers, we use an estimate (based on personal communication with Michael Brent) of 27 years old for all talkers. There is also no annotation of intonational phrase boundaries, which are known to affect word duration. However, in this corpus every pause of 300 milliseconds or more is taken to be an utterance boundary, so we use the utterance boundaries as a fairly robust approximation to intonational phrase boundaries. As in the ADS corpus, we remove all prosodic phrases which are three words or shorter, resulting in the corpus statistics shown in Table 1.

Pooled dataset To facilitate direct comparison between predictability effects in ADS and IDS, we created a pooled dataset containing the data from both `swbdnxt` and `Large Brent`. As can be seen in Table 1, the pooled dataset is relatively balanced in terms of the number of sentences and words from each type of speech, but not in terms of talkers. The imbalance in the number of talkers is handled by including a random effect for Talker in our model (described below).

Models

Approach Word duration is affected by many factors other than word predictability, such as talker age, speech rate, the word's length in phones, and its position in the intonational phrase. To control for these kinds of factors, we adopt a simple two-step regression procedure. First, we build a single control model (using the model selection procedure described shortly), which regresses log word duration against only control terms such as those above.[1] The control model is fitted to the pooled dataset, and includes a Speech Type term (ADS or IDS) to allow for non-predictability effects on duration due to speech type. We also allow interactions between Speech Type and the other control factors. Next, we take the residuals of this control model as the response variable for model selection among predictability terms, so that these terms can only be used to explain the part of the variance that has not already been accounted for by control factors. We perform three separate regressions on the residuals: one on the residuals from the ADS subset of the data, one on the IDS subset, and one on the entire pooled dataset. The ADS and IDS regressions allow us to assess which predictability terms are significant factors in predicting word duration in ADS or IDS, while the pooled regression can show, through interactions with the Speech Type term, whether a given predictability factor has different effects in ADS and IDS.

This two-step approach has the advantage that we do not need to worry about collinearity among our control predictors. If two predictors are collinear, the parameters of the terms in the control model will be unstable, but the overall variation explained, and the residuals, will be stable.

The approach outlined above could be used with a standard linear regression model. However, such a model assumes that data points are sampled independently, which is clearly not true for words from the same sentence or from the same talker. Instead, we use a mixed effects regression, which generalizes standard regression by including multiple random effects rather than only one (the error term). Specifically, we will be able to estimate different random baselines and slopes for each talker and for each sentence. The random effects will not be examined directly; rather, they will "soak up" the otherwise unhandled correlations between items, providing us with more robust parameter estimates without sacrificing statistical power. We discuss the details of the random effects and model selection procedure below, after describing the fixed effects used in our control and predictability models.

Fixed Effects: Control terms We include nine terms in the control model, based on those of (Bell et al., 2009). These are: **Talker Age** (taken from metadata in `swbdnxt`; estimated as described in the Data section for `Large Brent`), **Talker Sex**, **Speech Rate** (computed per utterance as $\log\left(\frac{\#\ \text{Vowels}}{\text{Second}}\right)$), **# Vowels** (taken from annotation based on a pronunciation dictionary), **Average Word Duration** (computed as the sum of the average duration of each phone in the word, following Bell et al. (2009); average phone durations were computed separately for each dataset), **Intonational Phrase Initial** (indicates whether a word is at the beginning of an intonational phrase; phrases are bounded by break indices of 3 or 4 in `swbdnxt` and are assumed to coincide with utterances in `Large Brent`), **Intonational Phrase Final** (as previous), **Content or Function Word**[2] (based on POS tags), and **Speech Type** (ADS or IDS).

Fixed Effects: Predictability terms[3] We include four predictability terms, again following Bell et al. (2009): **Log Word Frequency**, **Preceding Context** (log probability of a word given the preceding two words), **Following Context** (log probability of a word given the following two words), and **First Mention** (whether or not the talker has said the word). The first three predictors are all Good-Turing smoothed. To reduce collinearity, we residualized Word Frequency against the other three predictors.

Model Selection

We employ a model selection procedure that closely follows an algorithm introduced by Coco and Keller (2010), with only minor modifications to avoid specific interactions that lead to convergence errors (all our IDS talkers are female, for example, so we avoid testing for an interaction between Speech Type and Sex).[4] For each round of model selection, we consider two random effects: Talker, and Sentence (nested within Talker). We first determine which of the two random effects (intercept only) produces a better initial model. We then determine whether adding the other random effect (intercept only) produces a significant improvement in model fit. This is followed by another series of model comparisons to add fixed main effects and random slopes for each random effect that

[1] Like Bell et al. (2009), we take the log of the duration to avoid equating a 50ms difference in a word that is usually 60ms with a 50ms difference in a word that is usually 300ms.

[2] Bell et al. (2009) investigate this term in interaction with predictability terms. We attempted to include it in our predictability model, but it is highly collinear with other predictability terms and we failed to reduce collinearity to an interpretable level, so we include it in the control model instead.

[3] All of these terms are computed from the conjunction of the IDS and ADS corpora prior to discarding short prosodic phrases. We also tried computing these terms on the final dataset, after discarding short prosodic phrases, and found no significant differences in the modeling results.

[4] The R implementation of the modified algorithm is available at http://homepages.inf.ed.ac.uk/s0930006/modelselect.R.

is already in the model. Finally, we perform another series of model comparisons to add interactions.

In each step, a predictor is added if the model with that predictor is a significantly better fit to the data than the model without that predictor, as assessed with the `anova` function for model comparisons in `R`. As the model selection procedure involves several dozen model comparisons, we are conservative in assessing the significance of model comparisons. Specifically, whereas Coco and Keller (2010) consider a model comparison significant if the larger model is a better fit at $p < 0.05$, we require $p < 0.01$.

All predictors were centered except Speech Type; for ease of interpretation, Speech Type was set to -1 for Adult Directed Speech and 1 for Infant Directed Speech, resulting in a mean value of ≈ -0.071. During model search for the Predictability Models, we compel all terms to be added as at least a main effect, even if they do not improve model fit.[5]

P-values for fixed effects in all models (i.e., the P-mcmc values in Tables 2, 3, and 4) are assessed using a Markov Chain Monte Carlo algorithm, using $10,000$ samples, implemented in the `pvals.fnc` function from the R package `languageR`. A fixed effect is taken to be significant at $p < 0.01$ (the same results are found at $p < 0.05$).

Experiments

ADS and IDS individually

Before performing a direct comparison of ADS and IDS data, we first build individual predictability models for the two types of speech. The individual models serve two purposes. First, they tell us what kinds of effects are present in each kind of speech, allowing an informal comparison of the predictability effects in ADS and IDS speech. Second, we can use the results of these individual models to inform model selection when performing a direct comparison on the pooled data. In short, the individual models identify patterns of significant effects, and the pooled model compares these patterns of significant effects in a quantitative manner.

For the individual models, we take the residuals from the control model fitted on the entire pooled dataset, and run model search on the IDS predictability terms to predict the residuals for IDS words, and then run model search on the ADS predictability terms to predict the residuals for ADS words. Model search proceeds for up to 2-way interactions.

Results for ADS and IDS are presented in Tables 2 and 3, respectively. As the response variable is (residual) log duration, a negative coefficient corresponds to a shortening effect as the predictor increases, and a positive coefficient corresponds to a lengthening effect. For both ADS and IDS, a number of predictability effects are observed. For ADS, we find all the expected predictability effects among the main effects: frequent words and words predictable from context are shorter, while words new to the conversation are longer. Accordingly, we have replicated previous findings on ADS. In IDS, we also find effects, in the expected directions, of Word Frequency, First Mention, and Following Context, but we do not find any effect of Preceding Context. We discuss the implications of this difference in the Discussion Section.

Table 2: Coefficients and significance values (P-mcmc) for all fixed effects in the individual model for ADS data.

Predictability Term	Coeff.	P-mcmc
(Intercept)	-0.0045	0.2558
First Mention	0.0383	0.0001
Word Freq.	-0.0077	0.0042
Prec. Context	-0.0226	0.0001
Foll. Context	-0.0105	0.0001
Word Freq. $\times$ First Mention	-0.0339	0.0001
Prec. Context $\times$ First Mention	0.0132	0.0014

Table 3: Coefficients and significance values (P-mcmc) for all fixed effects in the individual model for IDS data.

Predictability Term	Coeff.	P-mcmc
(Intercept)	0.0035	0.3912
First Mention	0.0352	0.0001
Word Freq.	-0.0211	0.0001
Prec. Context	-0.0020	0.2490
Foll. Context	-0.0093	0.0001
Word Freq. $\times$ First Mention	-0.0206	0.0002

We find one significant interaction common to both ADS and IDS: a negative interaction between Word Frequency and First Mention, indicating that the lengthening effect of First Mention is reduced for more frequent words. The additional interaction between Preceding Context and First Mention in the ADS model is interesting, as its absence in the CDS model further supports the idea that Preceding Context influences word duration only in ADS.

These individual models, however, only reveal whether we have enough evidence to determine that the various coefficients are significantly different from zero, without comparing the effects in ADS with those in IDS. A direct comparison accomplishes two goals. First, it can confirm that the effect of Preceding Context is actually weaker in IDS than it is in ADS. Second, it is possible that an effect might be significant and in the same direction in both types of speech, but be much stronger in one type of speech. A pooled model on the entire dataset enables just such a direct comparison of effects in each Speech Type by examining interactions with the Speech Type term. We now proceed to this pooled model.

Pooled comparison

In this section, we perform model search on the full pooled dataset, and include in the predictability model the fixed effect "Speech Type" that indicates whether each word is from the ADS dataset or the IDS dataset. We will in particular be examining the interaction terms between Speech Type and the predictability terms. Since we wish to verify different

[5]In practice, all terms were significant in every case except for Speech Type. This is expected, as Speech Type was included in the control model.

patterns of significant results in models containing 2-way interactions between predictability terms, we perform model search up to 3-way interactions. Model search proceeds as before, except we force the addition of 3-way interactions between Speech Type and the two-way interactions which were added to the individual models.[6]

Table 4 presents the model coefficients and P-values from our final model. The table is separated into three boxes for each order of interaction, and each box is split into interactions not involving Speech Type on the top and interactions involving Speech Type on the bottom. There is relatively little collinearity in this model; most of the pairwise correlations among main effects are less than 0.1 in magnitude, and only First Mention and Following Context, along with Speech Type and First Mention, are above 0.2 (but below 0.3). Four correlations involving interaction terms are between 0.3 and 0.4 in magnitude, and five are between 0.2 and 0.3, but otherwise all interaction term correlations are less than 0.2, with the majority less than 0.1.

As our Speech Type variable is approximately centered, the main effects terms indicate an approximate average over the entire pooled corpus (with a slight bias to ADS). We observe in the main effects the same predictability effects as we saw in the ADS individual model: a significant and lengthening effect of First Mention, and a significant and shortening effect of Word Frequency and contextual probabilities.

Looking to the 2-way interactions that involve Speech Type, we first see a significant and positive interaction between Speech Type and Preceding Context, indicating that the shortening effect of Preceding Context is weaker in IDS. Moreover, since the interaction coefficient is roughly equal in magnitude to the main effect coefficient, this confirms the individual model findings that there is little or no effect in IDS. Secondly, we find a significant negative interaction between Word Frequency and Speech Type, indicating a stronger shortening effect of Word Frequency in IDS.

Two of the three 3-way interactions that involve Speech Type are significant. We see first a significant negative interaction between Speech Type, First Mention, and Preceding Context. As Speech Type is -1 for ADS, this verifies the discovery of a significant positive interaction between First Mention and Preceding Context in the individual ADS model but not in the individual IDS model. We find also a significant positive 3-way interaction between Speech Type, Following Context, and First Mention. This indicates that, for ADS, the lengthening effect of First Mention is diminished for words that are highly predictable in Following Context, while for IDS, the lengthening effect of First Mention is actually enhanced for words that are highly predictable in Following Context. Finally, the interaction between Speech Type, Word Frequency, and First Mention is non-significant, consistent with the similarly valued negative coefficients for the interaction between Word Frequency and First Mention in the individual models.

[6]We also tried forcing the addition of interactions between Speech Type and any term which was added to the individual models, but the additional interactions were not significant.

Discussion

In our experiments, we found many similarities in the effects of predictability on word duration in IDS and ADS, but also some significant differences. Specifically, we found significant main effects of First Mention, Word Frequency, and Following Context in both IDS and ADS, but of Preceding Context only in ADS. Moreover, IDS exhibits a stronger effect of Word Frequency. Also, while we found a significant negative interaction between Preceding Context and First Mention in ADS, the interaction was not significant in IDS. Finally, we found that words that are more predictable based on following context show a greater lengthening effect of First Mention in IDS, but a smaller lengthening effect in ADS.

These results are important for several reasons. First, we have confirmed the independent effects of First Mention, Frequency, and Preceding and Following Context found by (Bell et al., 2009), but using more stringent statistical methods which control for correlations due to Talker and Sentence, and exercising greater care with collinearity in our model terms. Second, we have provided the first detailed characterization of predictability effects in IDS, which adds to our knowledge about the nature of the linguistic input that infants are exposed to. Finally, we have added an important piece of evidence to the debate about whether predictability effects are listener- or talker-based: since ADS and IDS have the same population of talkers but different populations of listeners, differences in predictability effects between ADS and IDS suggest that these effects are, at least in part, modulated in response to listener characteristics. Our results cannot be explained under an information-theoretic approach by simply assuming a noisier channel in IDS, with talkers slowing down the overall rate of communication. Under this assumption, we would expect the same factors to be significant in predicting duration for both IDS and ADS. The coefficients might be different, but they would all change in the same direction. Instead we found heightened effects of some factors in IDS and lessened effects of others.

Our results also raise an interesting question: what could explain the particular pattern of differences we found? Although a fully satisfactory account requires further investigation, we speculate that the factors found to be significant in both ADS and IDS are those that reflect talker-based mechanisms, while the remaining factor (Preceding Context) reflects accommodation to the listener. Our reasoning is as follows. First, there is a well-attested relationship between word frequency and the talker-based process of lexical access (Griffin & Bock, 1998, and references therein); this relationship is usually explained in terms of the resting activation of particular lexical items. Similarly, talker-based lexical access processes can account for the effect of First Mention: previously mentioned words have higher activation due to priming. The effect of Following Context can be explained using a

Table 4: Coefficients and significance values (P-mcmc) for all fixed effects in the pooled model.

	Predictability Term	Coeff.	P-mcmc
Main effects	(Intercept)	-0.0012	0.6426
	Speech Type	0.0052	0.0636
	First Mention	0.0362	0.0001
	Word Freq.	-0.0145	0.0001
	Prec. Context	-0.0121	0.0001
	Foll. Context	-0.0099	0.0001
2-way	Word Freq. × First Mention	-0.0271	0.0001
	Word Freq. × Foll. Context	-0.0012	0.1612
	Speech Type × Prec. Context	0.0096	0.0001
	Speech Type × Word Freq.	-0.0065	0.0002
3-way	Word Freq. × Foll. Context × First Mention	-0.0046	0.0028
	Speech Type × Prec. Context × First Mention	-0.0099	0.0004
	Speech Type × Foll. Context × First Mention	0.0058	0.0072
	Speech Type × Word Freq. × First Mention	0.0070	0.0524

different talker-based mechanism, in this case sentence planning. This explanation arises from spreading activation models of sentence production (e.g. Dell, 1986; Tily et al., 2009), which tie contextual probabilities to syntactic form activation.

Although the effect of Preceding Context may also be related to lexical access, we note that it is the only measure which always involves words the talker has just said (and the listener has just heard). Accordingly, it is the measure best-situated to capture tacit awareness of the listener's processing load, and so may be the only measure which captures primarily listener-based influences on word duration. If this is the case, then it may be that talkers have no control over the first three effects (as they are hard-wired side effects of how language production works), but can control the effect of Preceding Context. Since infant listeners have little or no linguistic competence, talkers may simply "turn off" the effect of Preceding Context in IDS, knowing that the listener will be unable to make correct predictions about words in context.

Whether or not this explanation is correct, our results provide important evidence that talkers do modulate the effects of predictability on pronunciation based on coarse knowledge about listener characteristics. Further research is needed to better understand the extent and nature of this modulation, especially with other listener populations.

Acknowledgements: We would like to thank Frank Keller and Moreno Coco for invaluable feedback and comments.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31 – 56.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1-22.

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*, 92 – 111.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, 31 – 44.

Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., et al. (2010). The nxt-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, *44*(4), 387 – 419.

Coco, M. I., & Keller, F. (2010). Scan pattern in visual scenes predict sentence production. In *Proc. of the 32nd annual conference of the cognitive science society.*

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proc. of CogSci* (pp. 933–938).

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313–338.

Lindblom, B. (1990). *Explaining phonetic variation: A sketch of the H & H theory.* Kluwer Academic Publishers.

Rytting, C. A., Brew, C., & Fosler-Lussier, E. (2010). Segmenting words from natural speech: subsegmental variation in segmental cues. *Journal of Child Language*, *37*(3), 513 – 543.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2).