Bell et al. (2003)

Presenter: Joey Kurek

# Effects of disfluencies, predictability, and utterance position on word form variation in English conversation

# Overview



**Introduction & Methodology**

**Control Factors**

**Variables of Interest**

1. Disfluencies
2. Predictability
3. Word Position

**Conclusions**

# Part 1:

# Introduction

# Overarching Goal



Take advantage of online labeled corpora to explore **large prosodic, lexical, and environmental contexts** in natural speech production

Contextual variables in this study:

1. Presence of neighboring disfluencies
2. Word predictability from lexical context
3. Word position within an utterance

# Motivation/Background



## Disfluencies

Fox Tree and Clark (1997) showing that *the* is pronounced with a full vowel more often around disfluencies → Goal: **Expand findings to other lexical items**

## Predictability

Lots of work on role in access/disambiguation of structures but **not much work on role in production**

"**Probabilistic Reduction Hypothesis**" based on past studies: word forms are reduced when they have a higher probability (Jurafsky et al., 2001; Gregory et al., 1999)

## Prosodic Position

Three demonstrated effects:

- Final lengthening
- Initial strengthening
- Final weakening

Now specifically focusing on **natural speech**

Teasing apart prepausal lengthening from lengthening at the edge of prosodic domains

# Measuring Effects



## Dependent Variables

1. Duration of entire word
2. Categories of vowel quality
3. Presence of coda consonants

All reflect a scale of lenition, weakening, or reduction

## Lexical Scope

Ten frequent function words:

| | | | | |
|---|---|---|---|---|
| I | and | the | that | a |
| you | to | of | it | in |

Reasons for limitation:

1. More would be too ambitious
2. High frequency, high form variation, and monosyllabic
3. Typically not accented

**Key point for future research:** Do these effects hold for the rest of the lexicon?

# Vowel Coding



**Basic:**
citation/clarification pronunciation

**Reduced:**
[ə] [ə˞] [ɨ] [ɵ]

**Full:**
All except reduced

**Two binary variables:**
1. All full vs. reduced
2. Basic vs. other full

TABLE I. Most frequent pronunciations of the ten words, grouped into basic, full, and reduced-vowel pronunciations. For each word the three most common tokens of each type of pronunciation are listed in order of frequency.

| | Basic | Other full | Reduced |
|---|---|---|---|
| a | [eɪ] | [ʌ],[ɪ] | [ə],[ɨ] |
| the | [ði],[i],[di] | [ðʌ],[ðɪ],[ʌ] | [ðə],[ðɨ],[ə] |
| in | [ɪn],[ɪ],[ɪɾ], | [ɛn],[ʌn],[æn] | [ɨn],[ɳ],[ən] |
| of | [ʌv],[ʌ],[ʌvv] | [ɪ],[i],[ɑ] | [ə],[əv],[əf] |
| to | [tu],[tʉ],[ɾu] | [tʊ],[tɪ],[tʌ] | [tə],[tɨ],[ə] |
| and | [æn],[ænd],[æɾ] | [ɛn],[ɪn],[ʌn] | [ɨn],[ɳ],[ən] |
| that | [ðæ],[ðæt],[æ] | [ðɛ],[ðɛt],[ðɛɾ] | [ðɨt],[ðɨ],[ðɨɾ] |
| I | [aɪ] | [ɑ],[ʌ],[æ] | [ə] |
| it | [ɪ],[ɪt],[ɪɾ] | [ʊt],[ʊ],[ʌ] | [ɨ],[ə],[ət] |
| you | [yu],[u],[yʉ] | [yɪ],[ɪ],[i] | [yɨ],[y],[ɨ] |

# The Switchboard Corpus



- Collected by **Texas Instruments** in **1990**
- Used for development of speech recognition algorithms
- Contains a total of **240 hours** of speech coming from 2430 conversations between strangers
- Participants:
  - Paid volunteers
  - Native English speakers from all areas of the US
  - 543 total speakers (302 male, 241 female)
- Word-by-word transcription by court reporters

# Dataset for this Paper



**8362**
items for analysis

## Switchboard

**Audio with lexical transcription**

~3 million words

## ICSI

**Phonetic transcription**

~4 hour subset of original corpus hand-transcribed by linguistics students at UC Berkeley at International Computer Science Institute

## Treebank III

**Utterance segmentation**

1155 of the 2430 original conversations were segmented by the Linguistic Data Consortium (LDC) into about 205,000 utterance-like units

Two thirds of the ICSI corpus

# Regression Analysis



- A statistical model that predicts a response variable based on **contributions from a number of other explanatory factors**
  - Accounts for the many confounding factors of natural speech
- **Not actually used for predicting forms** in this study
- Word duration → ordinary linear regression
- Vowel quality/coda presence → logistic regression
  - Models **odds** of category = P(category) / [1 – P(category)]
- Important to look at effect size as well as level of significance

Potential for inflated significance values due to **non-independence of data points!**

Interpret results qualitatively

# Part 2:

# Control Factors

# Overview



Five control variables likely to play a large role based on prior research:

1. Rate of speech
2. Segmental context
3. Prosodic factors
4. Age/sex of speaker/hearer
5. Individual characteristics of function words

**Could be objects of study in their own right** (especially speaker/hearer effects) but here just treated as control factors

# 1. Rate of Speech



- Rate of speech affects all three measures of reduction (faster → more reduced)
- Limit effect for faster rates accounted for using both log(rate) and $\log^2$(rate)
- Shortening effects **not** simply a result of selecting reduced vowels
  - Rate still accounts for ~14% of variation after controlling for vowel reduction and coda deletion

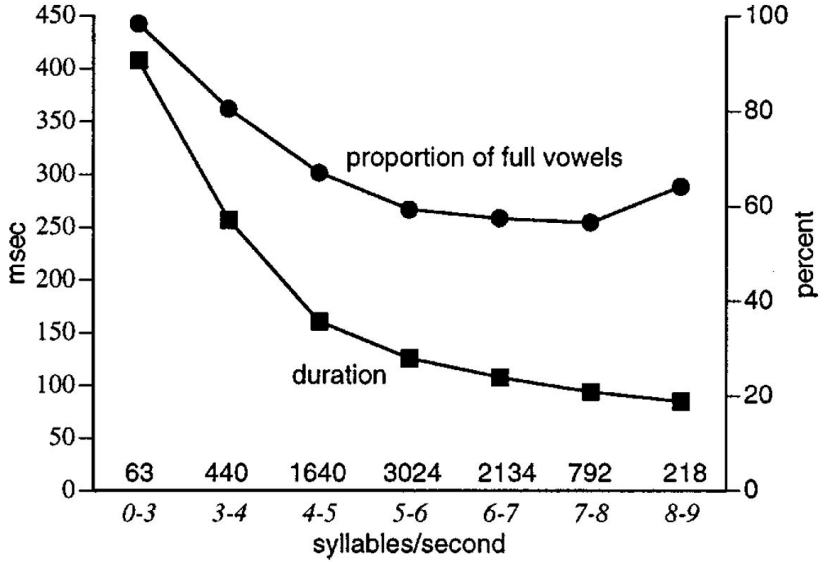



FIG. 1. Function word durations and proportions of full vowels by rate of speech. The scale for duration is on the left axis, the scale for full vowels is on the right. The number of observations for each rate category appears at the bottom of the graph.

# 2. Segmental Context



- Forms tend to be more reduced before consonants than before vowels (Rhodes, 1996, inter alia)
- Function words coded for whether vowel of following syllable is reduced/full
  - Bolinger (1986) predicts longer full vowels when following vowel is full → does not hold for this data
- Overall effects before consonant:
  - 1.63 times greater odds of reduced vowel
  - 0.79 shorter item durations
- As with rate, shortening effects **not** simply a result of selecting reduced vowels

TABLE II. Observed average durations and reduced vowel percentages of closed-syllable (VC) and open-syllable (V) function words before words beginning with consonants and with vowels.

| | Word | Next word | Duration (ms) | Percentage of reduced vowels |
|---|---|---|---|---|
| consonant | VC | CV | 132 | 33.7 |
| follows | V | CV | 102 | 45.6 |
| vowel | VC | V | 158 | 29.7 |
| follows | V | V | 128 | 33.0 |

# 3. Intonational Accent



- Desirable to include as response variable, but this entails analytic complexities
- Not transcribed in ISCI database, so no direct control or analysis of effects
- Hope that it does not affect overall analysis because function words not likely to be accented → verified by checking with small subcorpora of Switchboard

# 4. Speaker/Hearer Effects



- **Duration:** Words spoken by women and older speakers tend to be longer (gender gap more pronounced for older speakers)
- **Vowel Reduction:** Words spoken by men are more likely to be reduced
- **Coda Deletion:** Words spoken by women are more likely to have the coda deleted
- **Basic Vowel:** Basic vowels are used more by older speakers than younger speakers
- Listener not significant except more frequent reductions when speaking to younger listeners
- Agrees with general age/sex correlations

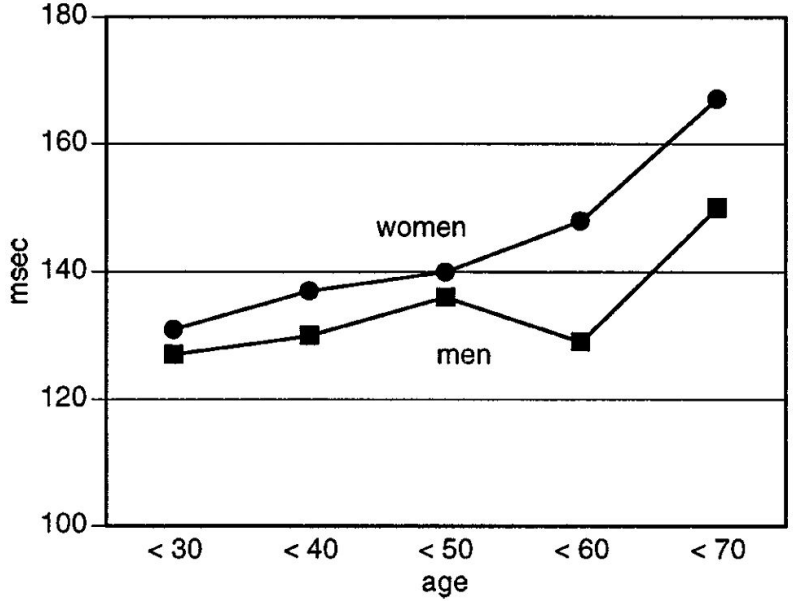



FIG. 2. Average word durations of function words of men and women speakers by age.

# 4. Speaker/Hearer Effects



- **General speech rate trends:**
  - **Men speak faster** than women
  - **Younger speakers** speak faster
  - Age/sex interaction: speech rate **gender gap more prominent for older speakers**

- **Disfluency trends:**
  - No gender gap when comparing uncontrolled averages
  - **Higher rate of disfluencies in men** when controlling for rate and probability variables

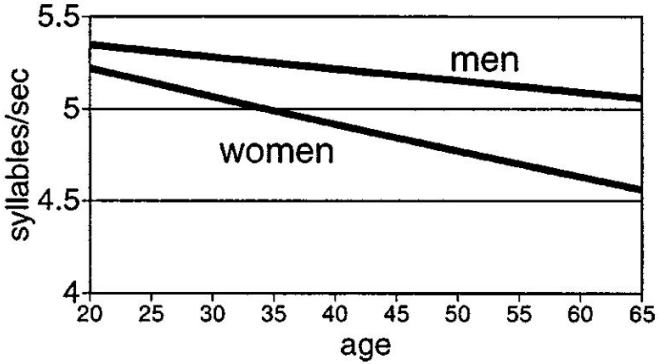



FIG. 3. Predicted average speech rates of men and women speakers by age.

# 5. Individual Characteristics



- Impractical to control for item effects in overall analyses
- Cannot investigate effects of lexical frequency
  - Narrow range of frequencies
  - Frequency confounded with other idiosyncrasies in small dataset
- Very different average durations
  - Conjunctions *and* and *that* notably longer, likely due to 1) intrinsically long vowel and 2) complex syllable structure

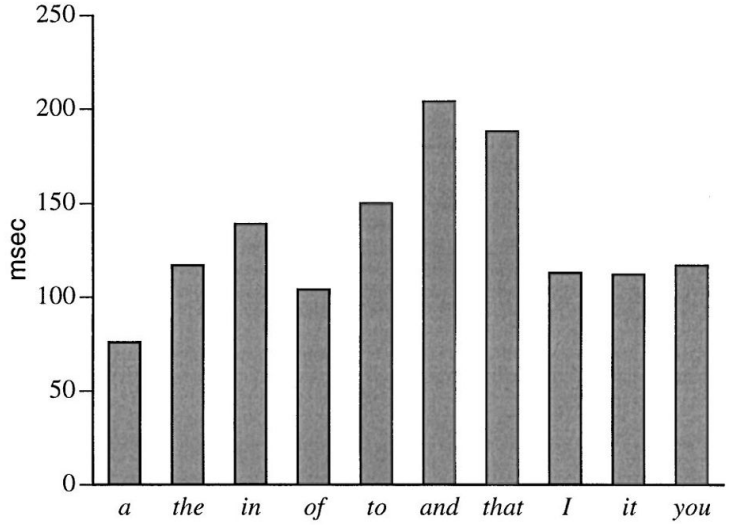



FIG. 4. Observed average durations of function words. The average duration of all words is 135 ms.

TABLE III. Frequencies of occurrence of the ten function words in the data.

| I | and | the | that | to | you | a | of | it | in | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 1381 | 1203 | 1123 | 786 | 769 | 758 | 745 | 583 | 562 | 452 | 8362 |

# 5. Individual Characteristics



- Rates of full vowels generally fall into high group (65–95%) and low group (25–35%)
  - Contributes to length differences
- Effects on results:
  - **Floor or ceiling effects**
  - Disproportionate representation of certain items in a certain context
- No direct control for item idiosyncrasies → just note consistency of effects over words

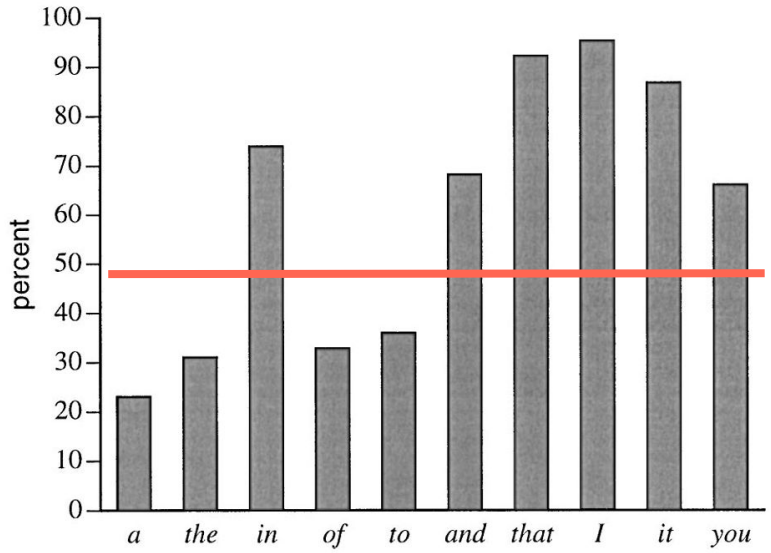



FIG. 5. Observed average frequency of occurrence of unreduced (full) vowels in function words. The average frequency for all words is 0.62.

# Part 3:

# Variables of Interest

# Disfluencies



Disfluencies arise when speakers have difficulty putting their thoughts into words and expressing them in proper syntactic and phonological form.

Past research links **longer and fuller forms** with planning problems. These forms could be:

1. Signals of planning problems
2. Production mechanisms to gain time

What counts as a disfluency? For this study:

1. **Silent pauses**
2. **Filled pauses**
3. **Repetitions**

| Following disfluency | Sentence |
|---|---|
| Repetition | **I** I have strong objections to that. |
| Silence | ...large numbers **of** (0.228) barefoot natives or something... |
| Filled pause (*uh*) | Somebody I talked to last week, they said they had **the** uh, they had problems doing some of the work |

| Preceding disfluency | Sentence |
|---|---|
| Repetition | I **I** have strong objections to that. |
| Silence | You know, the main things that I like about (0.214) **the** uh, job benefits... |
| Filled pause (*uh*) | it would encourage people, uh, **to** make more money |

# Effects of Disfluencies



Effects of a disfluent context after controlling for control/predictability factors:

1. **Duration:** Words are **1.34 times longer**
2. **Vowel reduction:** Words have **1.68 times greater** chance of having a full, unreduced vowel
3. **Basic vowel:** Odds of a basic vowel are **1.23 times greater**

(No coda deletion effect)

TABLE IV. Observed durations, frequencies of basic and full vowels, and frequencies of coda presence for function words in fluent and disfluent contexts. The number of observations of the context categories appears in parentheses. Basic vowel frequencies are based on the 4886 words with full vowels. Obstruent coda presence frequencies are based on the 2947 words *and, it, of,* and *that.*

| Context | Duration | Full vowel | Basic vowel | Coda presence |
|---|---|---|---|---|
| Fluent | 109 ms (5480) | 54% (5480) | 64% (2936) | 33% (1948) |
| Any disfluency | 187 ms (2519) | 77% (2519) | 64% (1950) | 39% (999) |
| Disfluency before | 137 ms (1295) | 73% (1295) | 59% (940) | 26% (366) |
| Disfluency after | 222 ms (927) | 80% (927) | 66% (741) | 42% (483) |
| Disfluency both | 295 ms (297) | 91% (297) | 75% (269) | 59% (150) |

# Location of Disfluencies



Disfluencies **more likely to occur** in the presence of other disfluencies.

The effects of disfluencies both before and after are **multiplicative**.

Following disfluencies have greater effect on duration. No difference for vowel reduction.

TABLE V. Occurrence of disfluencies before and after function words. The percentages of following disfluencies are also shown.

| | | Disfluency before | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Disfluency after | Yes | 297 | 927 | 1224 |
| | | 24% | 76% | 100% |
| | No | 1295 | 5480 | 6775 |
| | | 19% | 81% | 100% |
| | Total | 1592 | 6407 | 7999 |
| | | 23% | 77% | 100% |

TABLE VI. Estimated magnitudes and significance of the effects of disfluencies before and after a target word. The magnitudes for duration are the regression estimates of how much longer words are in the disfluent context. For the full vowel variable, they are estimates of the increase in the odds of occurrence of a full vowel in a disfluent context, compared to a fluent one.

| | Disfluency before | | Disfluency after | |
|---|---|---|---|---|
| Response variable | Effect | Significance | Effect | Significance |
| Duration | 1.22 | $F(1,6200) = 120.5,\ p < 0.0001$ | 1.51 | $F(1,6200) = 322.5,\ p < 0.0001$ |
| Full vowel | 1.59 | $\chi^2(1) = 27.8,\ p < 0.0001$ | 1.68 | $\chi^2(1) = 18.5,\ p < 0.0001$ |

# Interaction of Effects



Duration effects are **not** just a consequence of vowel reduction, nonbasic vowels, or coda deletion!

No interaction between presence/location of disfluency and vowel reduction. → **All vowels lengthened similarly.**

**All three disfluency types** have a lengthening effect.

# Item Effects

Impact of **syntactic class**: conjunctions/pronouns are located by disfluencies more than articles/prepositions

Conjunctions/pronouns also **longer** and **more frequent!**
→ Important to examine the effects of individual words

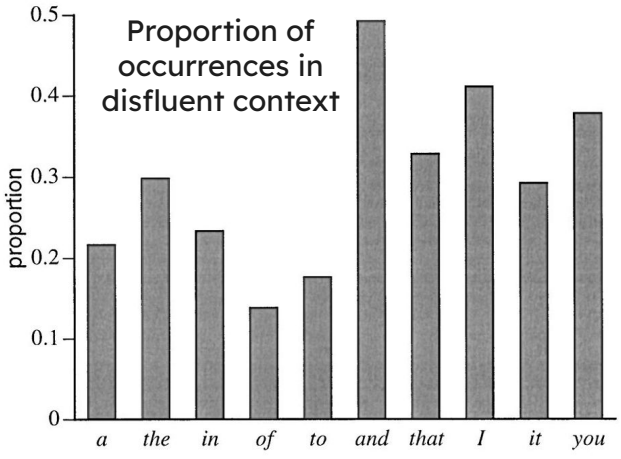



TABLE VII. Significances of the effects of neighboring disfluencies on individual function words. Preceding and following disfluencies have been collapsed for the vowel reduction and basic vowel variables.

| Effect on | *a* | *the* | *in* | *of* | *to* | *and* | *that* | *I* | *it* | *you* |
|---|---|---|---|---|---|---|---|---|---|---|
| Duration by a following disfluency | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| Duration by a preceding disfluency | <0.0001 | <0.0001 | ns | <0.005 | 0.01 | <0.0001 | 0.02 | <0.0001 | <0.0001 | ns |
| Reduced vowel by any disfluency | <0.0001 | <0.05 | 0.001 | 0.01 | <0.02 | <0.0001 | ns | ns | ns | ns |
| Basic vowel by any disfluency | ns | 0.02 | ns | ns | ns | <0.01 | ns | ns | <0.02 | ns |

# Word Predictability



Extension of *probabilistic reduction hypothesis*:
**more probable/predictable → more reduced**
(Jurafsky et al., 2001; Gregory et al., 1999)

Simplest measure: **prior probability** = relative frequency
- Not enough variation in this study
- Variation confounded with form/combination effects

**Joint probability**: prior probability of words taken together
- Can be high just because both words are frequent

**Conditional probability/mutual information** controls for the frequency of one or both words. For example, conditional probability of a word given the previous word:

$$P(w_i|w_{i-1}) = \frac{C(w_{i-1}w_i)}{C(w_{i-1})}.$$

TABLE IX. Summary of probabilistic measures and high probability examples.

| Measure | Definition | Examples |
|---|---|---|
| Joint of target with next word | $p(w_i w_{i+1})$ | **you** *know*, **I** *think* |
| Joint of target with previous | $p(w_{i-1} w_i)$ | *and* **I**, *in* **the** |
| Conditional of target given previous | $p(w_i \mid w_{i-1})$ | *rid* **of**, *kind* **of** |
| Conditional of target given next | $p(w_i \mid w_{i+1})$ | **I** *do*, **you** *know* |
| Conditional of target given surrounding | $p(w_i \mid w_{i-1} \cdots w_{i+1})$ | *matter* **of** *fact* |

# Duration Effects



Strongest effects from **conditional probabilities**

**Opposite effect** direction of higher joint probability depending on if it's with previous word (shorter) vs. next word (longer)

**Strictly local effects**

**Note:**
Effect ratios compare values at the 5th and 95th percentile of each variable

| Predictability variable | Duration | | | Full vowel proportion | | |
|---|---|---|---|---|---|---|
| | Significance | | | Significance | | |
| | $F$ | $p$ | Effect | $\chi^2(1)$ | $p$ | Effect |
| Conditional of target given previous | 88.4 | <0.0001 | 0.80 | 92.9 | <0.0001 | 0.24 |
| Joint of target with previous | 43.7 | <0.0001 | 0.94 | 55.2 | <0.0001 | 2.44 |
| Previous conditional×joint interaction | 58.7 | <0.0001 | | 20.4 | <0.0001 | |
| Conditional of target given next | 186.0 | <0.0001 | 0.72 | 22.3 | <0.0001 | 0.27 |
| Joint of target with next | 41.6 | <0.0001 | 1.20 | 272.8 | <0.0001 | 5.39 |
| Conditional of target given surrounding | 20.4 | <0.0001 | 0.91 | 2.9 | 0.09 | |

# Vowel Reduction Effects



Chances of full vowel based on conditional probability:

- 5th percentile → 0.73/0.72 (with word after/before)
- 95th percentile → 0.43

Higher joint probability **in either direction** is associated with less reduction → counterbalances conditional probabilities

Vowel effects **not** the only source of duration effects

**Note:**
Effect ratios compare values at the 5th and 95th percentile of each variable

| Predictability variable | Duration | | | Full vowel proportion | | |
|---|---|---|---|---|---|---|
| | | Significance | | | Significance | |
| | $F$ | $p$ | Effect | $\chi^2(1)$ | $p$ | Effect |
| Conditional of target given previous | 88.4 | <0.0001 | 0.80 | 92.9 | <0.0001 | 0.24 |
| Joint of target with previous | 43.7 | <0.0001 | 0.94 | 55.2 | <0.0001 | 2.44 |
| Previous conditional×joint interaction | 58.7 | <0.0001 | | 20.4 | <0.0001 | |
| Conditional of target given next | 186.0 | <0.0001 | 0.72 | 22.3 | <0.0001 | 0.27 |
| Joint of target with next | 41.6 | <0.0001 | 1.20 | 272.8 | <0.0001 | 5.39 |
| Conditional of target given surrounding | 20.4 | <0.0001 | 0.91 | 2.9 | 0.09 | |

# Item Effects



Each word affected by at least one predictability variable

Most general effect: conditional probability given next word

Varying predictability effects likely due to type of constructions that each function word appears in

- *And* in binomial constructions → bilateral cond. predictability
- *You know* → predictability effects similar after excluding

TABLE XI. Significances of the effects of predictability variables on individual function words. Effects with significances above 0.01 are in boldface.

| Effect on | a | the | in | of | to | and | that | I | it | you |
|---|---|---|---|---|---|---|---|---|---|---|
| Duration by conditional given following | **<0.05** | <0.001 | <0.0001 | ns | <0.0001 | ns | <0.0001 | ns | <0.005 | <0.0001 |
| Duration by conditional given previous | **0.05** | <0.001 | ns | ns | 0.0002 | ns | ns | ns | **<0.02** | ns |
| Duration by joint with previous | ns | **0.02** | ns | <0.0001 | <0.01 | ns | ns | ns | **<0.05** | ns |
| Reduced vowel by conditional given following | ns | ns | 0.0002 | ns | 0.0005 | ns | ns | ns | ns | <0.0001 |
| Reduced vowel by conditional given previous duration | ns | ns | **<0.05** | ns | ns | <0.0005 | ns | **<0.05** | ns | ns |

# Word Position



Three well-attested effects:

1. Final lengthening
2. Initial strengthening
3. Final weakening

Goals:

1. Move from lab paradigms to natural speech
2. Tease apart prepausal lengthening from lengthening at edge of prosodic domains

**Utterances:** sentence-like units making up conversation

- Complete syntactic sentences or phrases functioning as complete turns (e.g., "The news.")
- Possibly interrupted
- Compare: **turn**, **intonational/phonological phrase**

**Longer turn example**

**Utterance 1:** And, uh, I never really, messed with anything, uh, gardening or anything like that until now,

**Utterance 2:** but, uh, I, I keep hearing all the stories of, of different parts of town.

# Word Position Effects



Important: control for effects of disfluencies!

Initial words (vs. noninitial):

- **Longer** and **more likely to have full vowel**
- **No additional effect** if controlling for predictability.
  - Inherent characteristic of initial position

Final words (vs. medial):

- Also **longer** and **more likely to have full vowel**
- Effects resistant to following word predictability

| | Initial | Medial | Final |
|---|---|---|---|
| Duration (ms) | 173 | 125 | 200 |
| Vowel reduction | 82.3% | 57.4% | 93.2% |

# Word Effects & Discussion



Final lengthening more robust than initial lengthening. (Significant for all words vs. half of words examined.)

Initial position dominated by *and* (48%) & *I* (32%)

*And* exaggerates but does not single-handedly generate initial lengthening effect

Results confirm that reported **effects in laboratory settings extend to natural speech**

**Effects remain after controlling for final pauses**
→ prosodic edge lengthening ≠ prepausal lengthening

# Part 4:

# Conclusion

# Conclusion



"Disfluencies, predictability, and utterance position all play strong and independent roles in whether a word is reduced, for all measures of reduction."

## Disfluencies

Planning problems, both preceding and following, make words longer and less reduced

Some specific lexical items also show increases in basic vowel frequency

## Predictability

Function words are shorter and more reduced in highly predictable contexts (not limited to frequent collocations!)

Argument for cascade theory of word production: evidence for word selection passed on to lower levels of production

## Utterance Position

Words in utterance-initial and utterance-final position are longer and less reduced.

This is a result of the prosodic/syntactic boundary and not just because of the presence of pauses.

# Future Research




Apply these findings to **automated speech recognition**

- Sensitivity of models to speech rate and predictability
- Repetition detection
- Silence phone

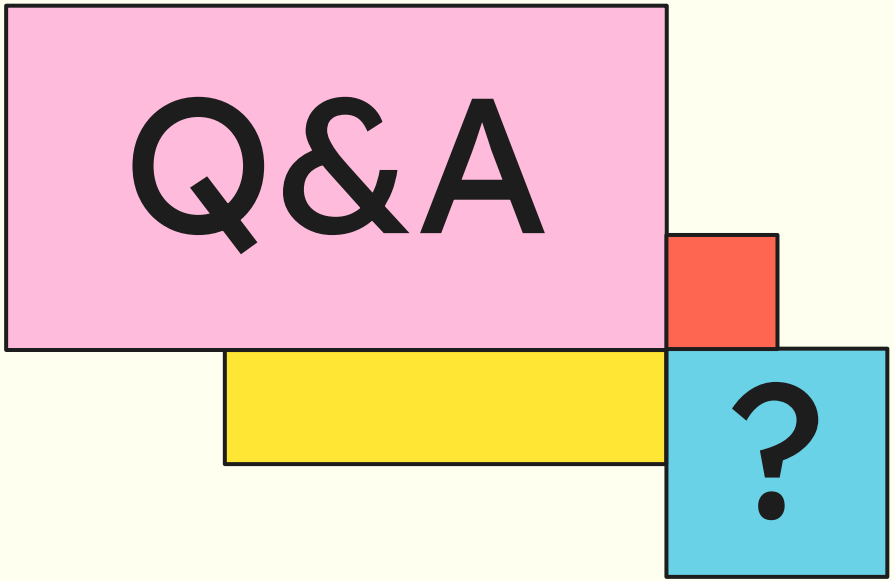
Examine to what extent these findings apply to **less frequent words**

Especially given:

- Fewer disfluencies
- Occur more freely

Take advantage of ecological validity of **corpus-based studies** (in conjunction with laboratory experiments)

- Trade-off with extensive coding and computations


Q&A
?