Expectation and Lexical Retrieval in Naturalistic and Experimental Misperception

Kevin Tang[1], Andrew Nevins[2]

[1]Yale University, Department of Linguistics & [2]University College London, Department of Linguistics; [1]kevin.tang@yale.edu, [2]a.nevins@ucl.ac.uk

Introduction and Data

- To understand aspects of the lexical retrieval system, we look into the errors created by the system from when it fails to retrieve the correct intended word.
- To increase the ecological validity of the findings, we examined misperceptions in both naturalistic and experimental settings.
- In this study, we systematically examined two naturalistic datasets collected in the "wild" (of conversational speech as well as sung speech) and one mega experimental dataset (of single-word lab speech).
- We examined the token frequency relationship between the intended word and the perceived word.

Word Confusion Data

- Conversational: 2,072 word confusion pairs
  - 1,231 monosyllabic words and 841 polysyllabic words
- Mondegreen: 9,609 word confusion pairs
  - 6,569 monosyllabic words and 3,040 polysyllabic words
- Experimental: 13,302 word confusion pairs
  - 5,385 monosyllabic words and 7,917 polysyllabic words

(Filtered pairs with function words, proper names, or different number of syllables.)

Corpus of Conversational Misperceptions

- A recently compiled corpus of naturalistic misperceptions in conversational speech by Tang (2015, Chapter 2).
- Consists of ≈ 5,000 misperception instances reported by individuals from six existing naturalistic speech misperception corpora of English

Corpus of Mondegreen Misperceptions (Misheard lyrics)

- KissThisGuy.com - The Archive of Misheard Song Lyrics
- Consists of instances of misheard lyrics reported by individuals
- 21,329 instances were extracted.

Corpus of Experimental Misperceptions

- A corpus of single word misperception (Felty et al., 2013)
- Tested 1428 words randomly sampled from the entire American English lexicon.
- The stimuli were masked with six-talker babbles at three signal to noise ratios (0, 5, 10dB), and were presented to 192 young, normal-hearing listeners in an open-set spoken word recognition task.
- 30,058 word misperception pairs were extracted.

Hypotheses and Predictions

Hypothesis 1 – Correlational

- When the intended word cannot be retrieved, listeners can **select a word that 1a) does or 1b) does not correlate with the intended word.**
- In a small experiment testing 144 words, Pollack, Rubenstein, and Decker (1960) analysed word frequency of the intended words and the perceived words and they did not find a significant correlation.
- In a larger experiment testing 1,428 words, Felty et al. (2013) found a positive correlation which is significant at a moderate level (R = 0.154, df = 21,842, p < 0.0001) between the intended and perceived words.

Hypothesis 2 – Frequency Differences

- When the intended word cannot be retrieved, listeners can **select a word that are 2a) more frequent, 2b) less frequent or 2c) neither more or less frequent than the intended word.**
- Vitevitch (2002), Bond (1999, p. 103) and Cutler and Butterfield (1992) examined a small sample (75 − 165) of naturalistic misperceptions. The differences between frequency of the intended and the perceived words were not significant.
- However, with experimental misperceptions, Felty et al. (2013) found that the perceived words have a higher frequency than their intended words.

Methods

Extractions of Word Confusions

- The misperceptions in the Conversational and Misperception corpora are of a sentence-level. Therefore, word confusions were extracted by relying on the mismatches between the phonetic transcription of the intended sentence and the perceived sentence.
- The phonetic alignments were achieved using a modified version of (Tang, 2015) of a PMI (Pointwise Mutual Information) based Levenshtein distance method (Wieling and Nerbonne, 2011).

Analyses

- To evaluate the stability of the results beyond the amount of segmental differences and monosyllabicity, the data were divided by 10 word-level Levenshtein distance bins as well as monosyllabicity.
- **Hypothesis 1: [Top-right]** Correlational analyses
- **Hypothesis 2: [Middle-right]** T-test analyses of frequency differences
- **Hypothesis 2: [Bottom-right]** To understand the discrepancy with the Mondegreen T-test results, the frequency distribution and vocabulary size were analysed for lyrics and everyday speech.

Conclusions and Implications

- Our frequency findings are stable against potential confounds such as monosyllabicity and the amount of word-level segmental differences.
- Our results suggest that listeners are sensitive to their experience and expectation of their phonetic and lexical knowledge during speech perception.
- Firstly, when contextual information (say, word duration) is available, listeners utilise them (as suggested by correlational findings in naturalistic data, and the lack of correlations in the experimental data).
- Secondly, listeners use their expectation of the frequency distribution of their lexicons. This is suggested by how the experimental data show that the listeners simply select high frequency words as the perceived words, since it is the best strategy in face of uncertainty due to signal degradation.

Furthermore, this is highlighted by the paradoxical result from the Mondegreen data of how listeners consistently *under*estimate the frequency of the intended words.

- Our work challenges research on single-word perception (Cooke, 2009; Tóth et al., 2015), especially on the mechanisms of lexical retrieval.
- It is clear the mechanisms of lexical retrieval function differently depending on the amount of contextual information provided.
- Contemporary models of lexical retrieval such as Shortlist and TRACE need to be able to account for these patterns of mis-retrieval.
- It further highlights the need for conducting misperception research beyond the word-level as well as the importance of ecological validity.

Data Availability

The conversational data used in this paper were extracted from the Slips of the EAR (SEAR) Project (www.searproject.org). It contains ≈ 5000 instances of speech perception error of English collected in everyday settings. The archive provides an interactive exploration interface.

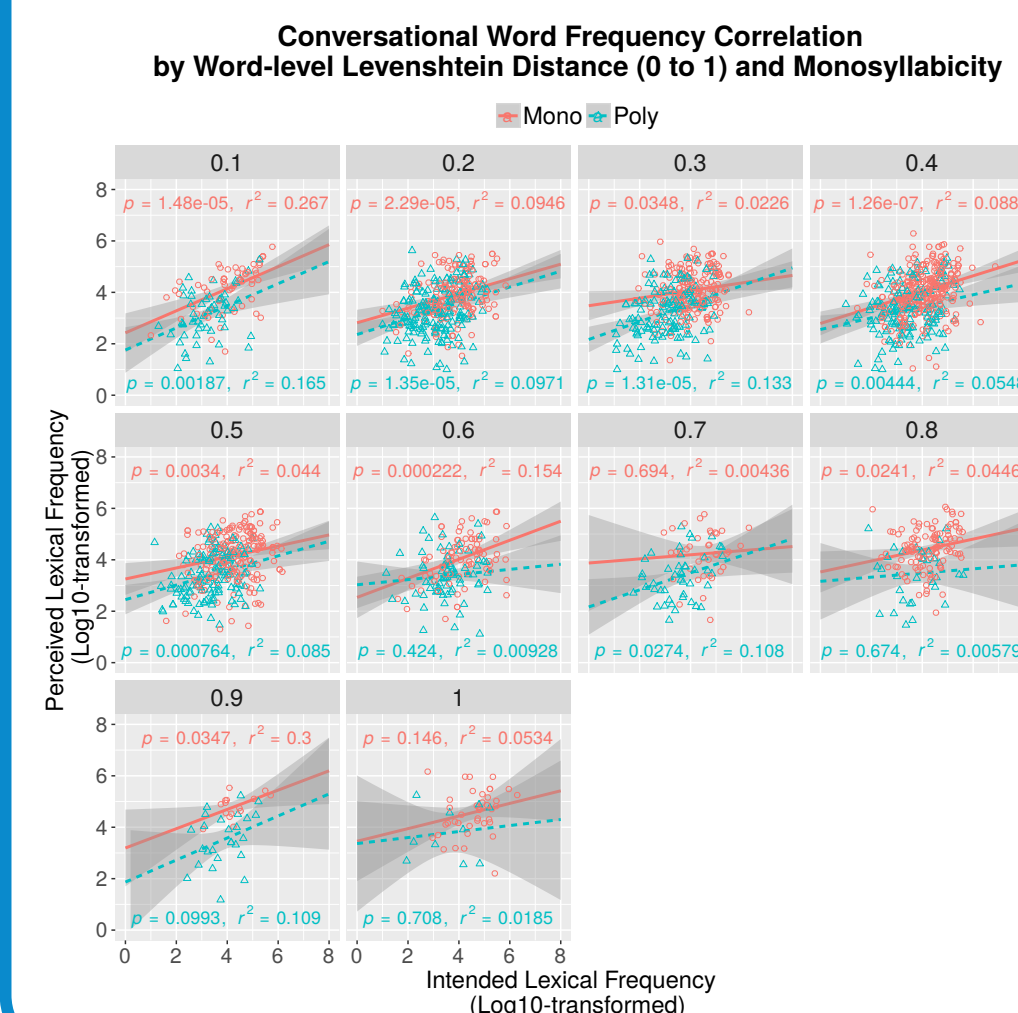## Caveats and Further Directions

- The frequency distribution of lyrics could be genre-dependent.
- Mondegreen instances might be more humourous than conversational instances, therefore they contain less frequent words.
- Will experimental *sentence* misperceptions be similar to naturalistic ones?
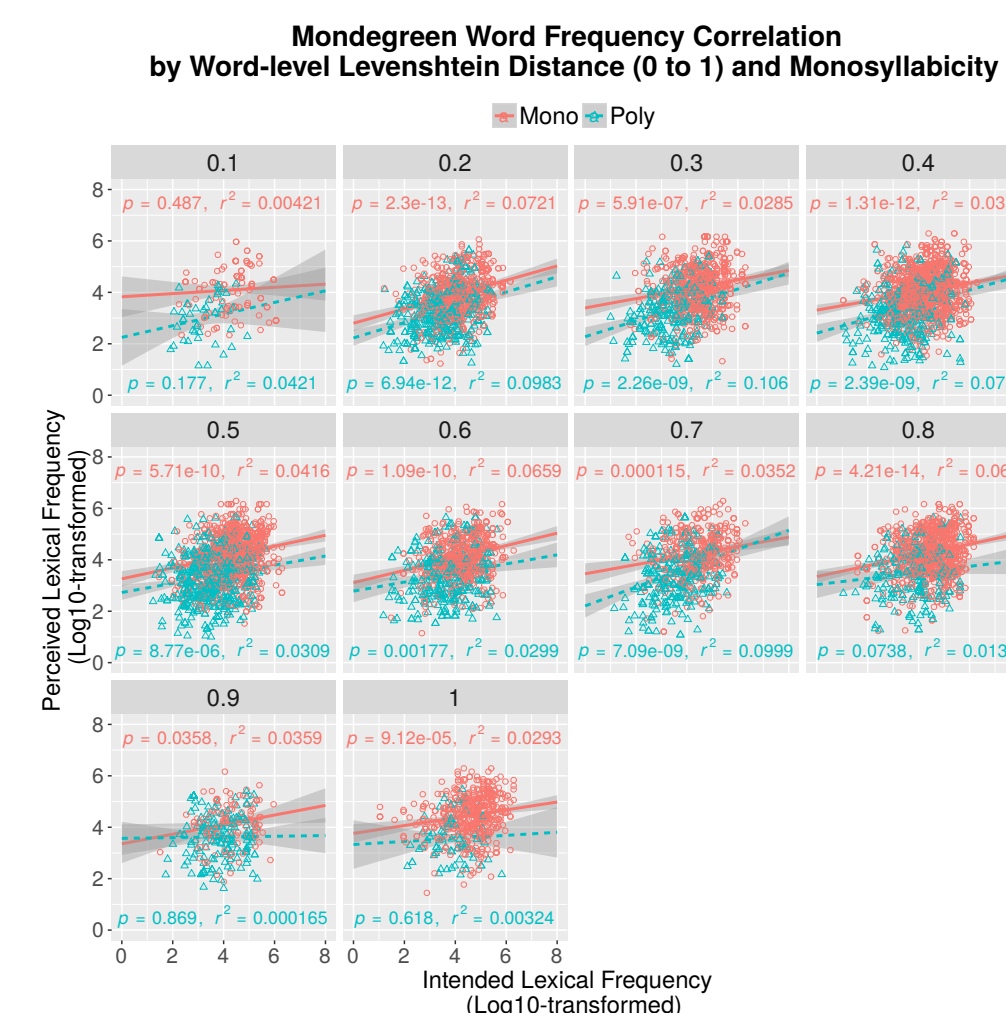
Selected references

Baayen, R.H. (2001). *Word frequency distributions.* Vol. 18. MIT Press.

Bond, Zinny S. (1999). *Slips of the ear: Errors in the perception of casual conversation.* New York: Academic Press.

Cooke, Martin (2009). "Discovering consistent word confusions in noise". In: *Proceedings of Interspeech.* Brighton, UK, pp. 1887–1890.

Cutler, Anne and Sally Butterfield (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception". In: *Journal of Memory and Language* 31.2, pp. 218–236. DOI: 10.1016/0749-596X(92)90012-H.

Felty, Robert Albert et al. (2013). "Misperceptions of spoken words: Data from a random sample of American English words". In: *The Journal of the Acoustical Society of America* 134.1, pp. 572–585. DOI: 10.1121/1.4809540.

McClelland, James L., David E. Rumelhart, and Geoffrey E. Hinton (1986). "The appeal of parallel distributed processing". In: *Parallel distributed processing: Explorations in the microstructure of cognition.* Ed. by David E. Rumelhart, James L. McClelland, and The PDP Research Group. Vol. 1. London: MIT Press, pp. 3–44.

Pollack, Irwin, Herbert Rubenstein, and Louis Decker (1960). "Analysis of incorrect responses to an unknown message set". In: *The Journal of the Acoustical Society of America* 32.4, pp. 454–457. DOI: 10.1121/1.1908097.

Tang, Kevin (2015). "Naturalistic Speech Misperception". PhD thesis. University College London.

Tóth, Máté Attila et al. (2015). "A corpus of noise-induced word misperceptions for Spanish". In: *The Journal of the Acoustical Society of America* 137.2, EL184–EL189. DOI: 10.1121/1.4905877. URL: http://dx.doi.org/10.1121/1.4905877.

Vitevitch, Michael S. (2002). "Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear". In: *Language and Speech* 45.4, pp. 407–434. DOI: 10.1177/00238309020450040501.

Wieling, Martijn and John Nerbonne (2011). "Measuring linguistic variation commensurably". In: *Dialectologia: revista electrònica*, pp. 141–162.

Wright, Charles E. (1979). "Duration differences between rare and common words and their implications for the interpretation of word frequency effects". In: *Memory & Cognition* 7.6, pp. 411–419. DOI: 10.3758/BF03198257.

Is there a correlation between the frequencies of the perceived word and the intended word?
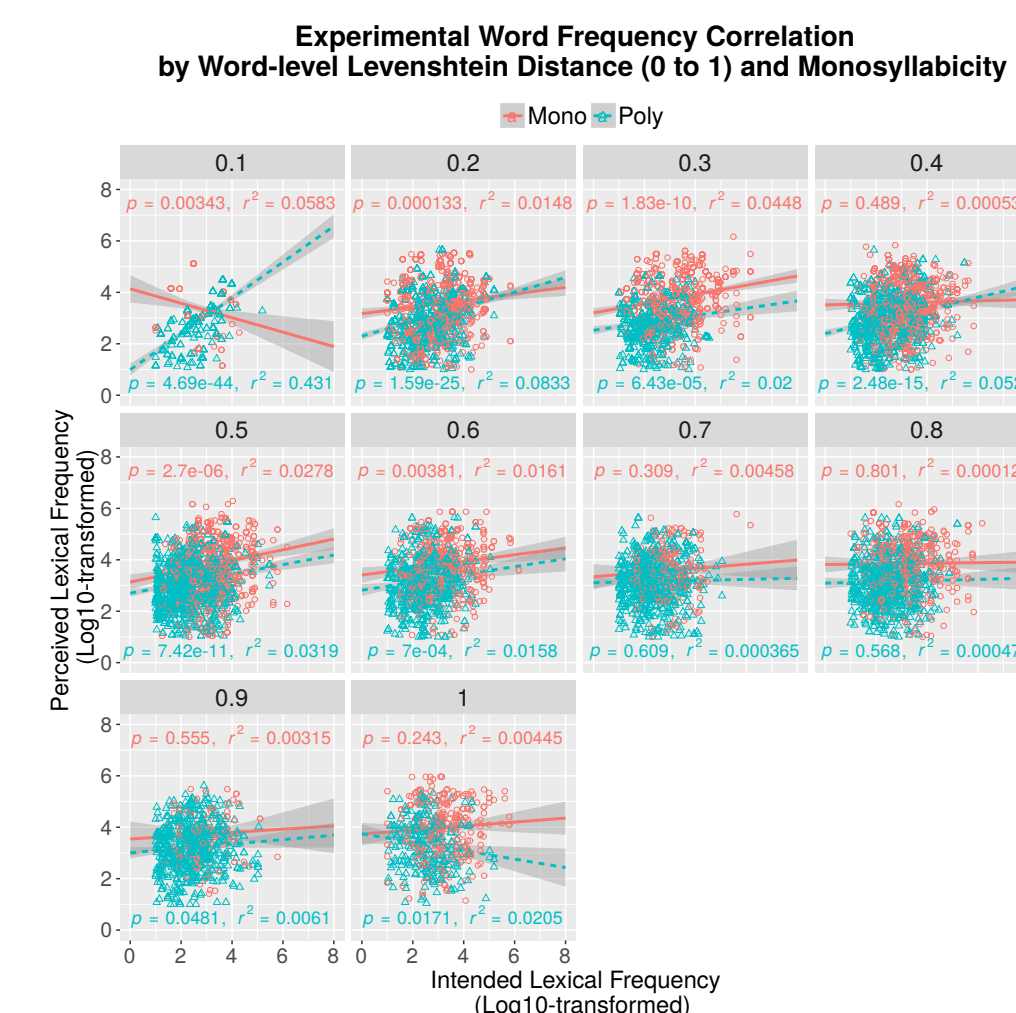
Conversational

Conversational Word Frequency Correlation by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

Mondegreen

Mondegreen Word Frequency Correlation by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

Experimental

Experimental Word Frequency Correlation by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

The figures show the correlation between the token frequency of the intended word (x-axis) and that of the perceived word (y-axis) for each word-level Levenshtein distance bin and for monosyllabic words and polysyllabic words separately. **Red:** monosyllabic words. **Blue:** polysyllabic words.

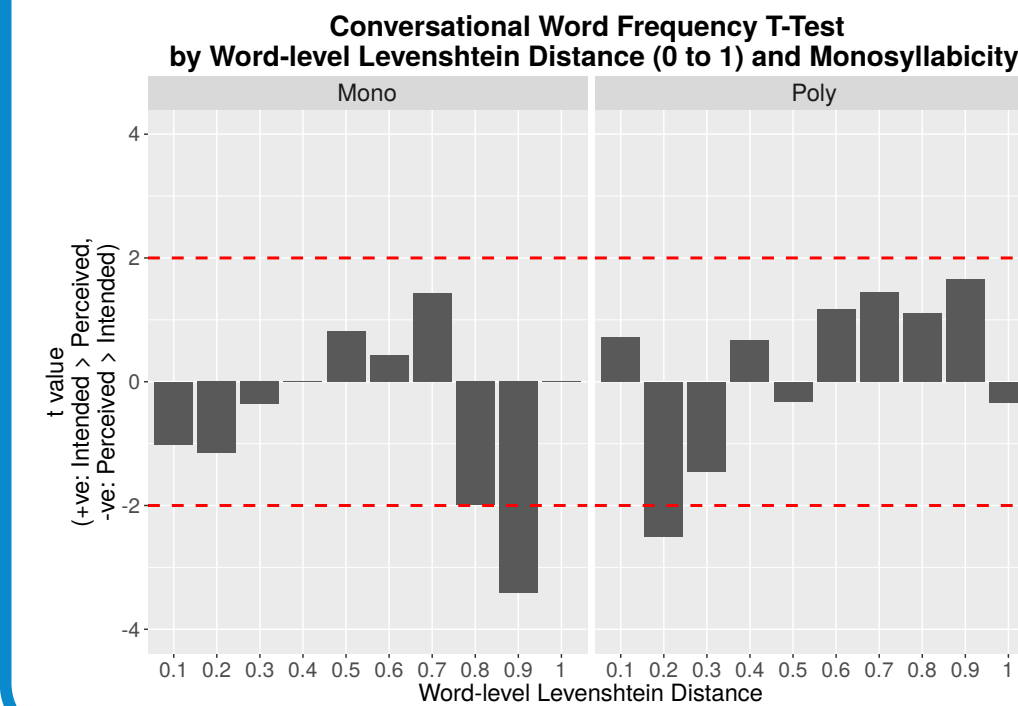### Summary of the Correlations

	Overall R^2		Across
	Mono.	Poly.	Bins
Convers.	8.5%	8.2%	Positive
Mond.	5%	5.5%	Positive
Exp.	1.3%	2.7%	Flat/Neg.

Is the frequency of the perceived word higher/lower than or the same as the intended word?
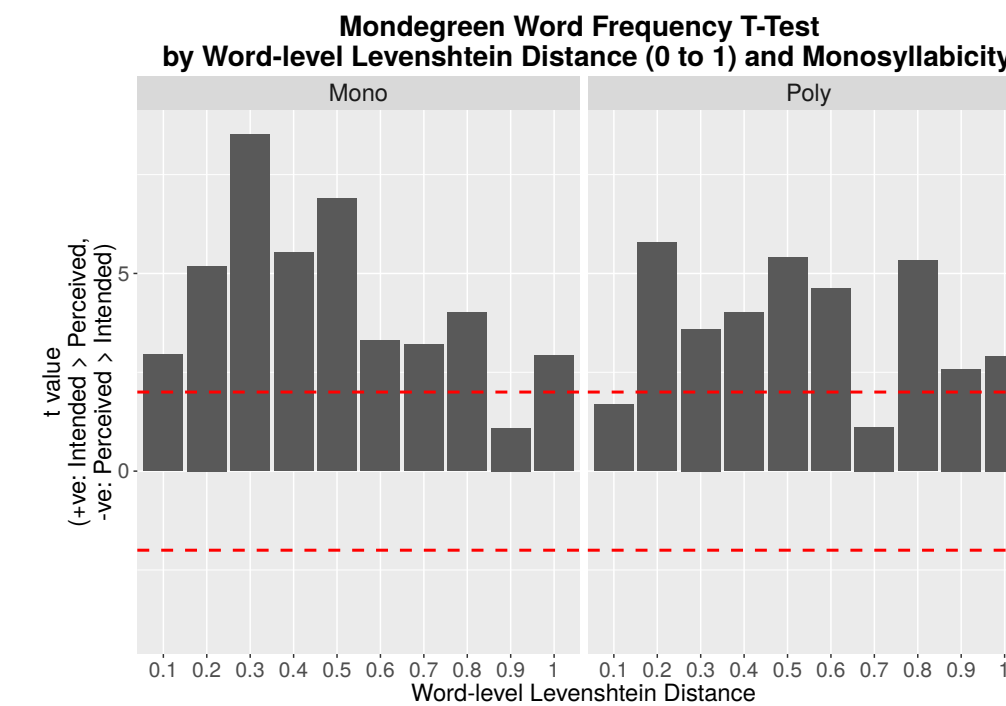
Conversational

$Freq._{Perceived} \approx Freq._{Intended}$

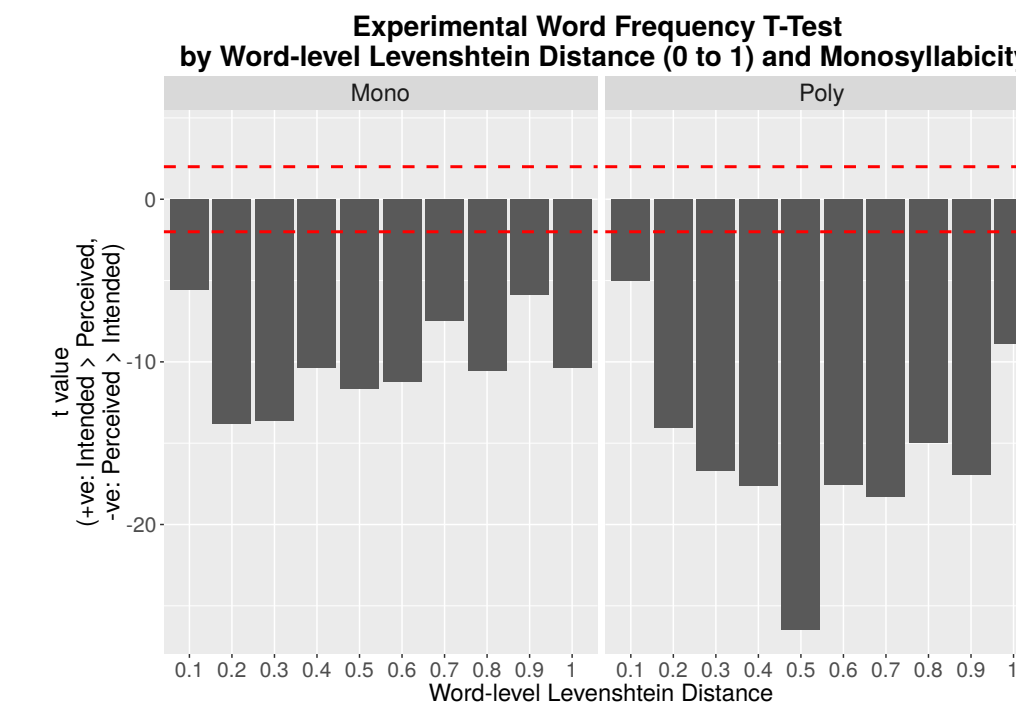Conversational Word Frequency T-Test by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

Mondegreen

$Freq._{Perceived} < Freq._{Intended}$

Mondegreen Word Frequency T-Test by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

Experimental

$Freq._{Perceived} > Freq._{Intended}$

Experimental Word Frequency T-Test by Word-level Levenshtein Distance (0 to 1) and Monosyllabicity

Experimental

$\uparrow \Delta Freq._{Diff} = \uparrow Noise$

Experimental Word Frequency T-Test by Noise Level (+10 - 0 dB) and Monosyllabicity
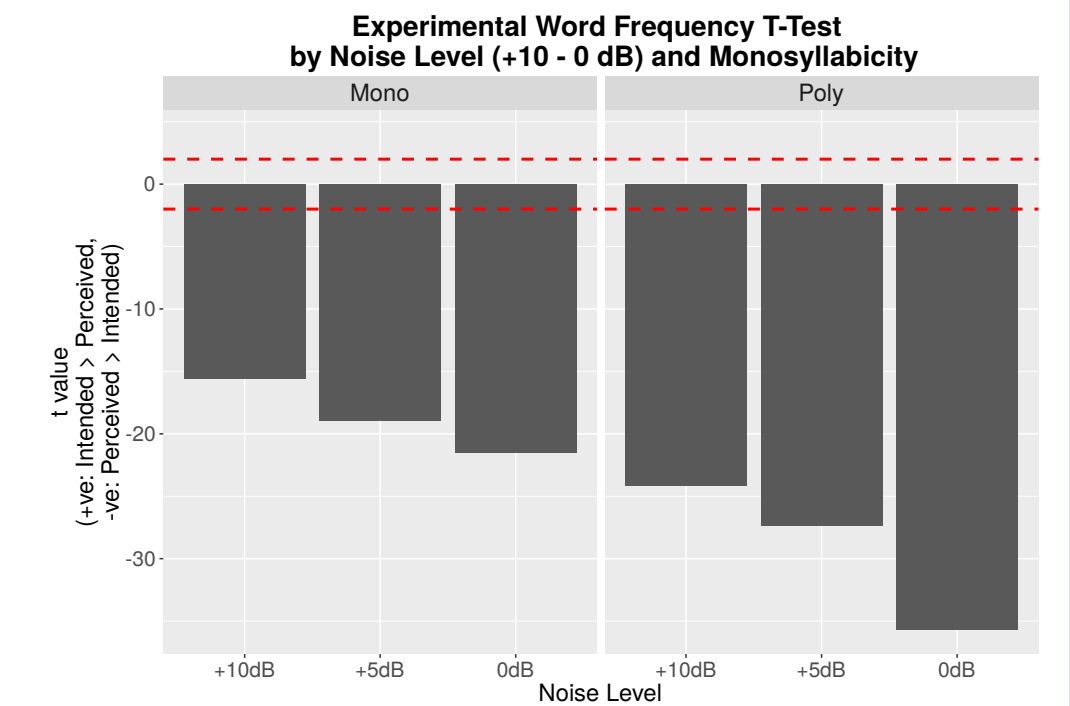
Frequency Expectation of Lyrics and Everyday speech

More high frequency words in Lyrics

Frequency Distribution Conversational vs. Mondegreen
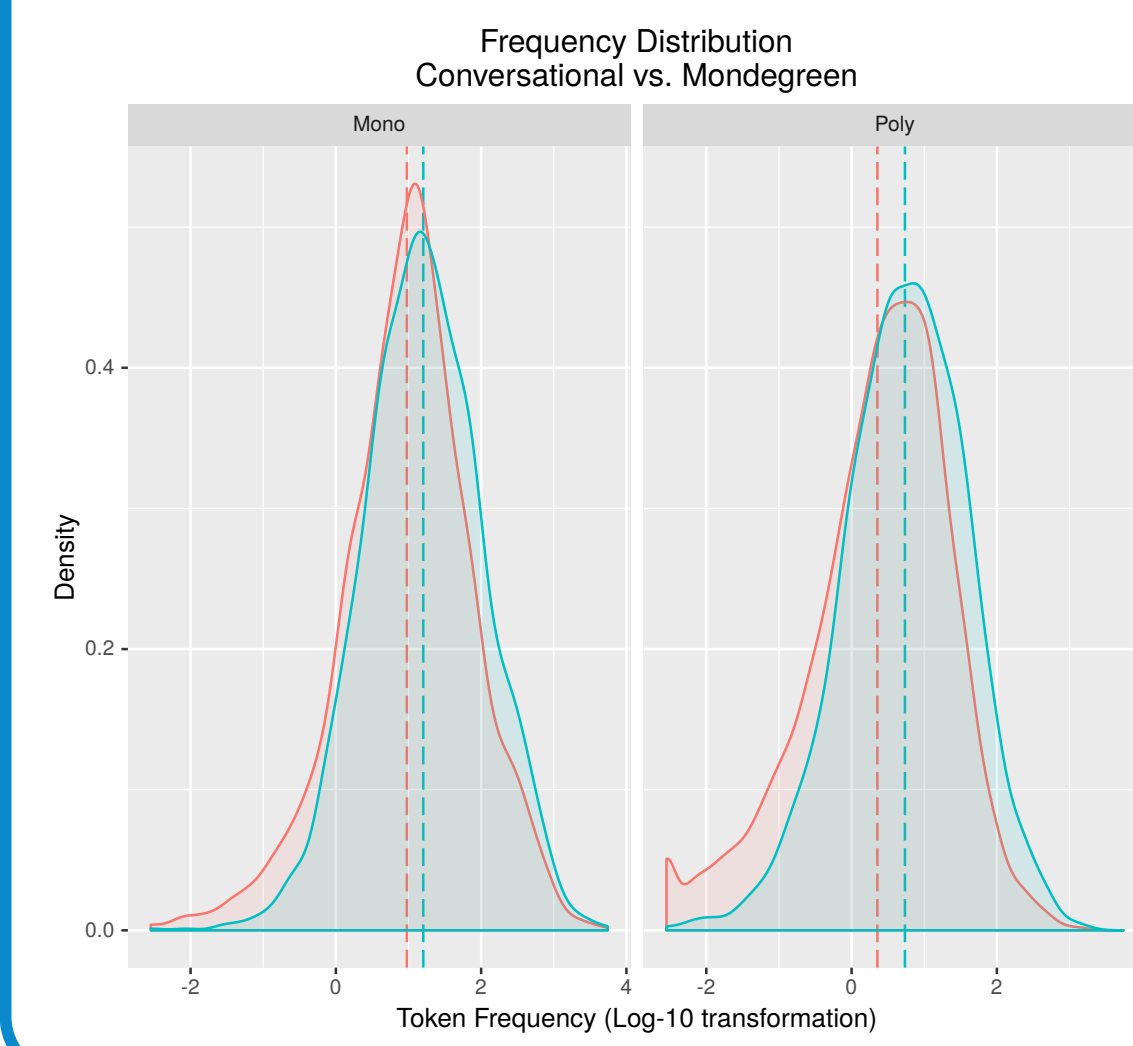
Vocabulary Size is smaller in Lyrics

Growth Curve of Type/Token Ratio by Monosyllabicity
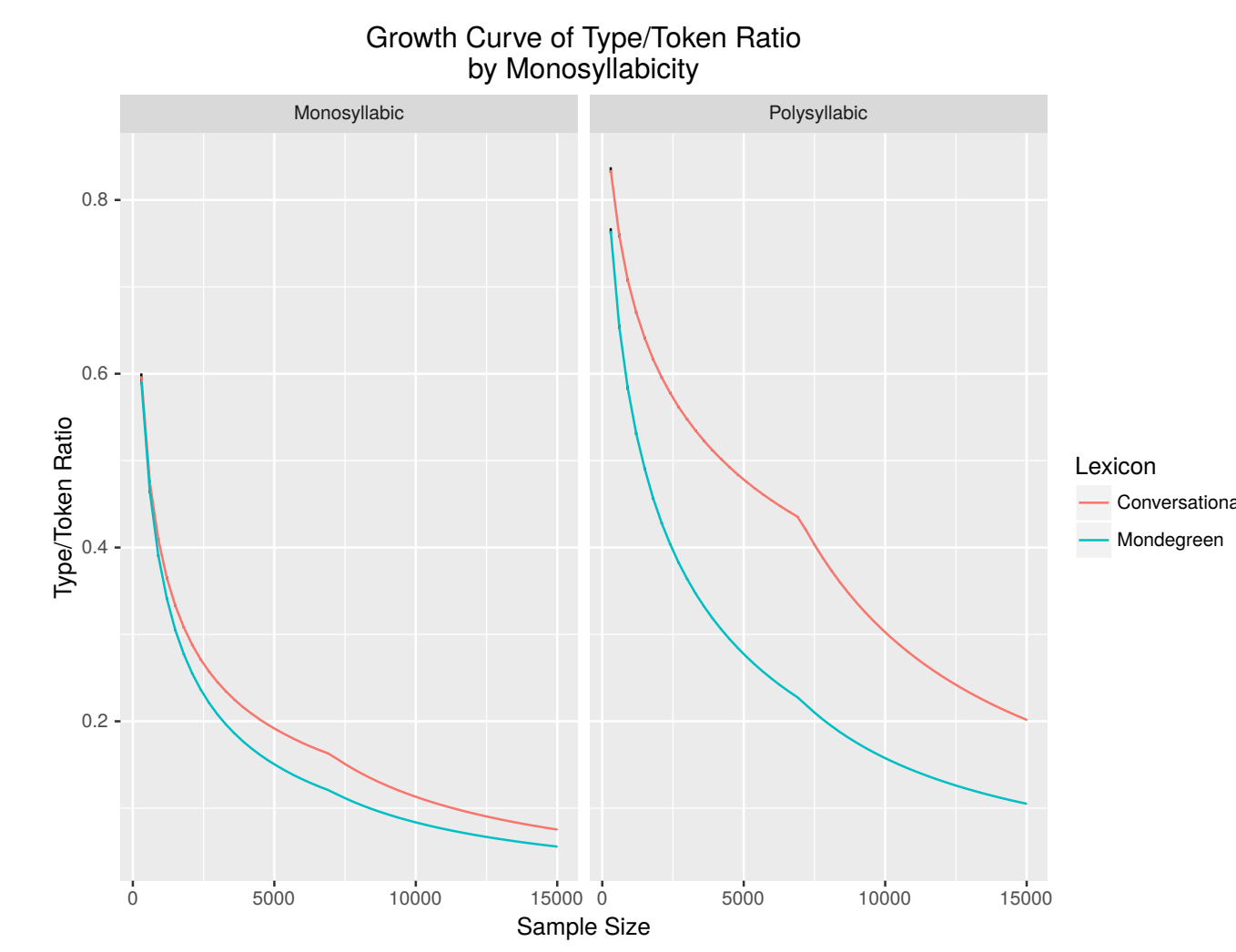
- In the Mondegreen data, hearers would perceive a less frequent word than the intended word
- The intended sentences of the Mondegreen data were used as a sample of lyrics (117,326 tokens), while subtitle texts were used as a sample of everyday speech (353 mil. tokens).
- **Skewness** Frequency distributions of 40,000 word random samples suggest that lyrics contains more high frequency words than everyday speech.
- **Vocab Size** Growth curve analyses (Baayen, 2001) with 15,000 word random samples showed that the vocabulary size (type/token ratio) is consistently smaller for lyrics than for everyday speech.

Discussion

Conversational vs. Mondegreen

- Naturalistic Mondegreen data exhibited a correlation between the intended word and the perceived word, but the paired T-tests showed that the perceived word is less frequent than the intended word.
- Hearers selected a word that is correlated with the intended word, which means the hearers are estimating the intended words using contextual information, but they consistently select a similar but less frequent word.
- This paradoxical result was addressed by the fact that there tends to be more frequent words in lyrics than in everyday speech, therefore the intended word would tend to be a more frequent word than expected on the basis of the frequency distribution of everyday speech.
- This fact is explicable by the fact that the vocabulary size is smaller for lyrics than for everyday speech, and this would explain the over sampling of high frequency words because there are simply fewer words (specifically low frequency words) to sample from.
- Assuming that hearers keep track of the frequency distribution of everyday speech and not of lyrics (at least less so), the result is that they would underestimate the frequency of the intended word.

Conversational vs. Experimental

- Conversational naturalistic data exhibited a correlation between the intended word and the perceived word and the paired T-tests showed that the perceived word is neither more or less frequent than the intended word.
- Experimental data exhibited no correlation and the perceived word tends to be more frequent.
- We attributed this to a graceful degradation account of lexical retrieval. Graceful degradation is the ability of a processing system to not break down in a catastrophic way when the input is incomplete, but to output a representation that best matches the input (McClelland, Rumelhart, and Hinton, 1986). In the context of misperception, the perceptual system uses the information in the degraded signal to retrieve a lexical item (Vitevitch, 2002). Naturalistic conversational data are rich in context, such as relative word duration, and syntactic, pragmatic and semantic information.
- High frequency words tend to be produced more quickly than low frequency words which tend to be produced more slowly (Wright, 1979). If listeners access a word with the same approximate duration as the intended word, they are likely to access a word with the same approximate frequency.
- In the experimental data, the words were presented in isolation, without relative word durational information or other contextual cues. Therefore, the hearers selected the most likely candidates in terms of the frequency distribution.