Massive reduction in conversational speech and its implications for auditory word recognition

A presentation based on Keith Johnsons paper "Massive reduction in conversational American English"(2004)

Presented by Adrian Bleymehl

What is "Massive Reduction"?

- "The phonetic realization of a word involves a large deviation from the citation form such that whole syllables are lost and/or a large proportion of the phones in the form are changed." - Johnson 2004

In the paper, "Massive Reduction" is described through the terms of

-Syllable Deletion

-Segmental Changes

What is "Massive Reduction"?

(1)	dəvɪnəti	fʌdʒ
	dəvĩɪ̯ti	fʌdʒ
	dəvĩ̯ĩ̯	fʌdʒ

- Example : A range of variants of the phrase " divinity fudge" by David Stampe(1973)

- Syllable Deletion: The most reduced variant has only two syllables, while the citation form has four.
- Segmental Changes: Only three of eight citation segments are preserved in the most reduced variant.

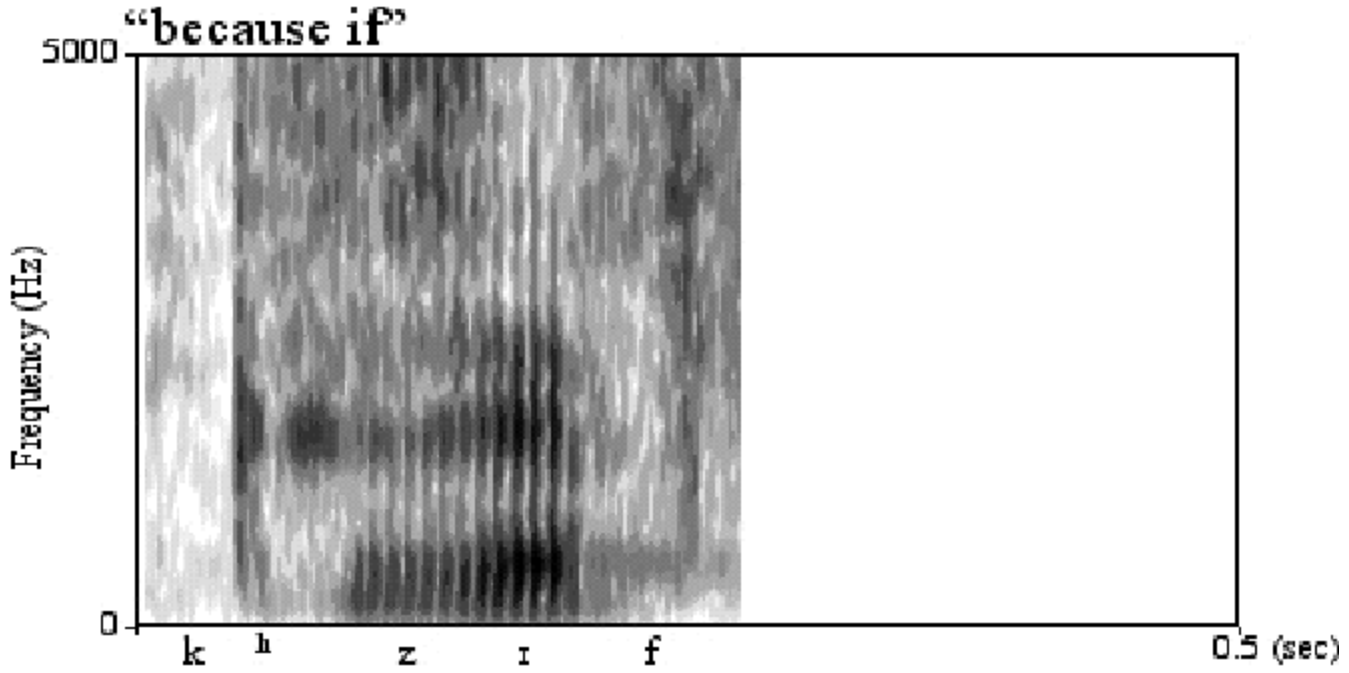
Implications of Massive reduction

"The goal of the paper is to relate pronunciation variation to models of auditory word recognition" (Johnson 2004)

- Researcher have only focused on segment-count preserving variants in auditory word recognition models . ---> Massive reduction does not necessarily preserve segment-count of words. (Syllable Deletion)

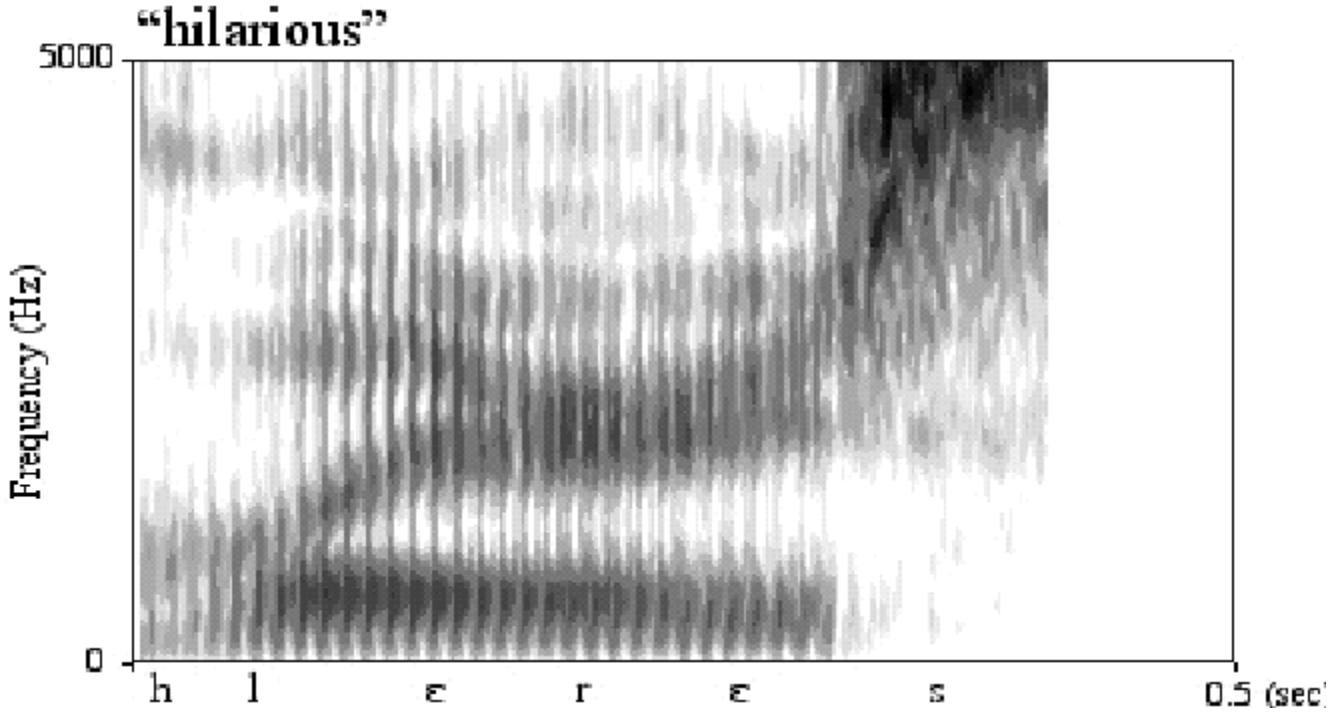
- If Massive reduction can be proven to be frequently in conversational speech, phone-by-phone segment-count preserving look up procedures will not be able to work.

More Examples of Massive reduction



- Zero-Syllable realisation of the two-syllable function word "because", forming an “illegal” consonant cluster.
- Only two out of five segments are retained in the new production.

More Examples of Massive reduction



- The four syllable word "hilarious" being reduced to a two syllable word. [hɪlɛrɪʌs] ~ [hlɛrɛs]

Massive reduction from a phonological standpoint

- Massive reduction is mostly irrelevant in the field of phonology, as many phonologists discard them as "vulgar" or "slovenly" pronunciations.

- Knott (1935) explains that editors of dictionaries often disregard the sounds of words in sentences and only deal with words spoken in isolation, often recorded from speech teachers.

Massive reduction from a phonological standpoint

- Hockett (1965) believes that a spectrum of speech styles exists, with "frequency norm" representing more normal rapid speech, and "clarity norm", which represents extra clear speech.

- Hockett remarkes that we tend to prefer the "frequency norm", but that this kind of speech style has "consequences". And if we don’t accept them, we try to avoid them by using "clarity norm".

Massive reduction from a phonological standpoint

- In general, most phonological theorists disregard highly reduced speech, as their goal is to study the information content of language sound systems.

- Johnson criticizes taking theoretical phonological analysis as the starting point for a theory of auditory word recognition, that tries to show the listener's ability to cope with phonetic variation.

Mass reduction from the standpoint of phonetics

- Phoneticians studying casual speech phenomena have consistently noted reductions in spoken language involving extreme change from citation forms.

- Cruttenden (1994), studying Old English, has noted that some common reductions in Old English changed to be the "established forms" of words in Modern English, regardless of speaking style.

- For example, words like "evening" or "camera", with two spoken syllables instead of the old three syllables, which are reflected in the spelling.

Mass reduction from the standpoint of phonetics

- Dalby (1986) did an extensive research on schwa deletion. In a study of "consultative" speech produced in television shows, he finds 9% of all schwas deleted, while the deletion rate in very fast speech was up to 43%.

- In his paper, Dalby (1986) suggests that "the overall strategy in “fast speech” is to reduce the number of syllables in an utterance while preserving the well-formedness of surface syllabification.”

Mass reduction from the standpoint of phonetics

(2) Reduced forms of *probably*, with a syllable count and number of unchanged phones for each form.

	prɑbəbli	prɑbli	prɑli	prɑˈ
σ	3	2	2	1
# unchanged phones	8/8	6/8	5/8	2/8

- An example of schwa deletion from Dalb (1986). You can see different variations of the word "probably" ranging from three to one syllable.

Study motivation

- Johnson (2004) wants to examine whether massive reduction occurs with any frequency in conversational speech. For this, he explores the "Variation of Conversation" corpus for massive reductions.

- After a frequency of massive reduction is found, Johnson wants to discuss the relevancy of these findings for different auditory word recognition theories.

The "Variation in Conversation" corpus

The "Variation in Conversation" corpus is a large database of recorded conversational speech:

- It consists of speech from a sample of 40 talkers, all native from Central Ohio.
- The sample is limited to middle class Caucasians and also stratified for age and sex.
- The talkers spoke a total of around 300.000 words, of which 88.000 words have been phonetically transcribed.
- At the moment of the study, the recordings of 14 of the 40 talkers had been phonetically tagged.

Measuring Massive reduction

To measure massive reduction in the corpus, Johnson conducts two different analysis :

- Analysis of syllable deletion: For this analysis, Johnson defined a set of syllable peaks. Then he counted and compared the syllable peaks in each citation form and their variation forms.

- Analysis of segmental deviation: In this analysis, Johnson counted the amount of segments in the variations, that deviated from the corresponding segments in the citation form.

Results: Syllable deletion

Content words:

- 2945 out of 38560 content word tokens had not the same number off syllables ---> a syllable mismatch rate of 7.6 %.
- 78% of those involved syllable deletion instead of syllable insertion, giving content words a 6% rate of syllable deletion.

Function words:

- 2592 out of 46774 function word tokens had not the same number of syllables --> a syllable mismatch rate of 5.5%
- 86% of those involved syllable deletion instead of syllable insertion, giving function words a 4.5% rate of syllable deletion.

Results: Segmental deviation

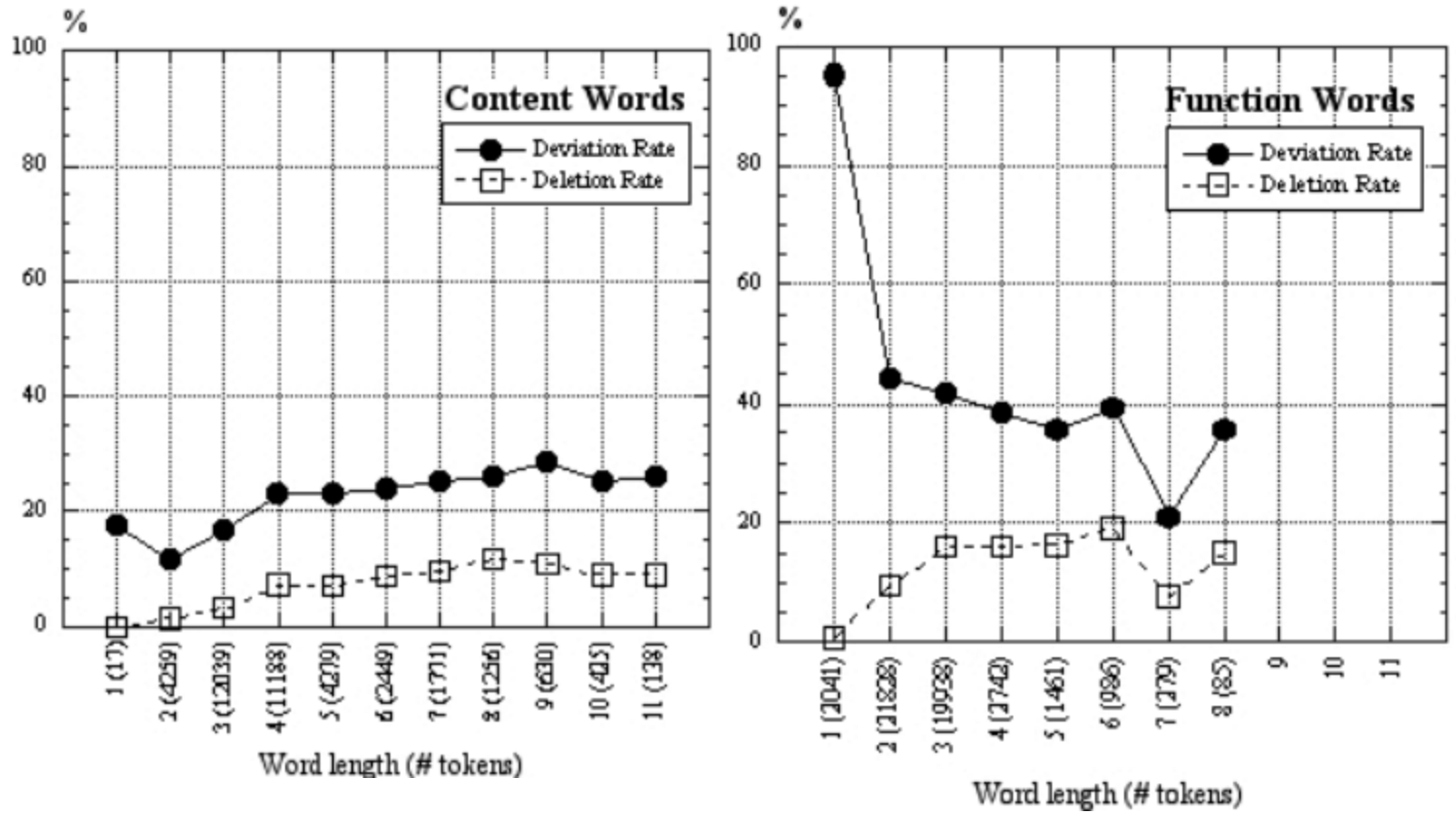
Content words:

- Content words had a deviation rate of <20% for words with three or less phones.
- Deviation rate of 20-25% for words longer than three phones.

Function words:

- Function words had deviation rates of around 40%.
- Exceptions for words with one or seven tokens, with around 90% and 20% deviation rates respectively.
- One-phone words included the indefinite article "a", which was rarely produced in its citation form. Seven phone words include "anybody", "somebody"…, words that showed rates of deletion and deviation more like content words.

Results: Comparing graphs



Deviation and deletion rates of content words vs function words

Results: Example

Table 2: Variants of *until* classified by the number of phones that deviation from the citation form and the number of phones that are deleted relative to the citation form.

Transcription	# deviating phones	# deleted phones
ʌntɪl	0	0
ʌntə̲l ɛ̲ntɪl	1	
ɛ̲ntə̲l ɪ̲ntɪw̲	2	
_n̩tɪl		1
ə̲ntʌ_	3	
_n̩t_l̩	4	2
__t l̩ __tə̲_		3

As it can be seen on this table, segmental deviation correlates with syllable deletion, as deleted segments will not be able to match the citation form.

Discussion

- The results of the study shows that massive reduction is indeed a regular feature in conversational speech.
- Johnson wants to focus on auditory word recognition theory and the implications that reduction has on its traditional dictionary assumptions.
- These traditional assumptions are the following:

(1) The segmental assumption : Lexical forms are composed of phonetic segments that are analogous to the letters of an alphabetic writing system.

(2) The single-entry assumption: Lexical forms are stored in a single prototypical representation.

Discussion

- Johnson wants to show that with the phenomenon of massive reduction, neither of these traditional assumptions about the formal representation of lexical items is tenable.

- To do this, we will now discuss four different types of auditory word recognition models :

-Single-entry, segmental models

-Single-entry, nonsegmental models

-Segmental, multiple-entry models

-Nonsegmental, multiple-entry models

(1) Single-entry, segmental models

- Most models of auditory word recognition fall into this category, including the Cohort model (Marslen-Wilson 1987) or the TRACE model (McClelland & Elman, 1986).

- All single-entry segmental models operate with a two step word recognition process.

- As single-entry models, they represent each lexical item with just one possible pronunciation in the mental lexicon.

(1) Single-entry, segmental models

- The requirement that segments in the input have to line up, left-to-right, with segments in the lexical representation, prevents single-entry segmental models from being able to cope with deletion.

- Therefore, these models predict that words with deleted segments or syllables cannot be identified by listeners.

- Johnson suggests that the input-to-lexicon mapping which has been assumed in these models needs to be revised.

(2) Single-entry, nonsegmental models

- The single-entry, nonsegmental model proposed by Lahiri(1999) claims that the human auditory word recognition system uses a featurally underspecified lexicon (FUL).

- By underspecifying certain features of the lexical representation, like "place of articulation", you can avoid mismatches in variations.

- "Feature underspecification is a clever method of soaking up variance by reducing the number of features that must be matched in lexical access"(Johnson 2004)

(2) Single-entry, nonsegmental models

- FUL has the ability to deal with massive reduction, as deleted segments/features score "non missmatches", if the lexicon doesn’t specify a contradictory feature.

- Non missmatch features are not penalized in the model, which leads to the problem that for example, the input [kæp] results in the cohort of equally viable candidates cap, camp, crap, clap, car wrap,...

- Johnson also notes other problems of the model :

(1) Phonological “inferencing” effects

(2) The featural unpredictability of deletion

(3) Subcategorial phonetic residue

(3) Multiple-entry, segmental models

- Pronunciation variation can be expressed in lists of alternate pronunciations using alphabetic symbols.

- By adding multiple-entry lexical representations to the originally single-entry model Shortlist, Scharenborg & Boves (2002) managed to drop the word error rate from 64% to 48%.

- As researchers have attempted to build recognizers that are capable of recognizing conversational speech, many have begun to explore multiple-entry lexica.

(3) Multiple-entry, segmental models

Johnson mentions two problems with these multiple-entry models:

- Segmental multiple-entry models introduce confusions into the lexicon. For example, the schwa-deletion variant of "ago" would be the same as the entry for the lexical word "go".

- Segmental multiple-entry models are still segmental.
  This includes a fixed set of phonetic symbols, which can lead to exaggeration of the lexical confusion problem.

(4) Multiple-entry, nonsegmental models

- Klatt (1979) proposed a model of speech perception, based on how well the acoustic/spectral shape of words matches stored finite-state models of possible spectral sequences representing words.

- X-MOD (Johnson 1997) is an extension of Klatt's model, which calculates an auditory spectral representation of incoming speech and sweeps this representation over an exemplar covering map.

Conclusion

- Through his experiment, Johnson (2004) proves that "massive" reduction does occur frequently in conversational speech.

- For this reason, models of auditory word recognition have to abandon their traditional "dictionary" assumptions about the auditory mental lexicon.

Sources

Main Source :

- https://www.coli.uni-saarland.de/courses/sprachproduktion/2024_WS/papers/johnson_massive-reduction_2004.pdf (Johnson (2004): "Massive reduction in conversational American English")