DISSIMILARITY THREE WAYS

BENGALI CORONALS SEEN THROUGH: ECHO REDUPLICATION CONFUSABILITY AND LEXICAL STATISTICS

SAMEER UD DOWLA KHAN, REED COLLEGE

IIT DELHI COLLOQUIUM, 18 NOVEMBER 2016

OUTLINE

- **Overview**
  - Identity and similarity in phonology
- **Echo reduplication**
  - Identity avoidance, with a puzzle from English
  - Production data from Bengali
  - Gradient similarity avoidance
- **Similarity metric**
  - Shared natural classes
  - Weighted shared natural classes
- **Other expressions of gradient similarity**
  - Lexical statistics
  - Perceptual confusability
- **Synthesis of results**

CATEGORICAL IDENTITY/SIMILARITY

- Many processes incorporate **categorical identity**
  - **Reduplication**
  - **Sibilant harmony** in Chumash (Beeler 1970)
    - [kiʃkín] + [us] ⇨ *[kiʃkínus] ⇨ [kiskínus] 'I saved it for him'
  - **Haplology** (identity avoidance)
    - merry + -ly$_{ADV}$ ⇨ merrily$_{ADV}$
    - silly + -ly$_{ADV}$ ⇨ *sillily$_{ADV}$ ⇨ silly$_{ADV}$
    - friend+ -ly$_{ADJ}$ + -ly$_{ADV}$ ⇨ *friendlily$_{ADV}$ ⇨ friendly$_{ADV}$

- Often broadened to a natural class: **categorical similarity**
  - **Vowel harmony** in Turkish (Tosun 1999)
    - [kɯz] + [ɬar] ⇨ [kɯzɬar] 'girls'
    - [jyz] + [ɬar] ⇨ *[jyzɬar] ⇨ [jyzlʲær] 'faces'

GRADIENT SIMILARITY

- But some phenomena in perception and the lexicon are best described as involving **gradient similarity**
  - Lexical **cooccurrence** effects in Muna (Coetzee & Pater 2005)
    - [d] is found in fewer roots with [t] than with [n]
  - Perceptual **confusability** in English (Cutler et al. 2004)
    - [tʃ] is misidentified as [t] more often than as [dʒ]
- It's possible that cases of supposed categorical identity/ similarity are in fact extreme cases of gradient similarity
  - cf. Vowel harmony in Hungarian (Hayes & Londe 2006)

GOAL FOR THIS TALK

- Present data illustrating the phenomenon of **fixed-segment echo reduplication** in (Bangladeshi) Bengali
- Demonstrate that it is a case of **gradient similarity**
- Explore what kind of **metric** underlies the patterns seen
- Investigate **lexical and perceptual** expressions of similarity as a comparison

ECHO REDUPLICATION

- **Echo reduplication**
  - Most common in lgs across southern Asia, e.g. Hindi:
  - [nɑm $ʊ_F$ɑm] 'name(s), etc.'
  - [roʈi $ʊ_F$oʈi] 'bread, etc.'

- Phonological properties
  - Total reduplication
  - Systematic replacement of some material in reduplicant (RED) with one or more **fixed segments**

- Semantic properties
  - Typically denotes **generalization**: 'X, etc.', 'superset of X'
  - In some lgs, it can also be **disparaging**

ECHO REDUPLICATION

- Representative examples:

- Fixed C in Turkish [m_F] (Southern 2005)
  - [kutu] 'box' ⇨ [k**u**tu **m_F**utu] 'box(es), etc.'
  - [ʃɑkɑ] 'fun' ⇨ [ʃɑkɑ **m_F**ɑkɑ] 'easily', 'calmly'

- Fixed V in Eastern A-Hmao [$ú_F$] (Mortensen 2006)
  - [ámâ] 'eye' ⇨ [ám**$ú_F$** âmâ] 'eyes, ears, mouth, and nose'
  - [kíl̥áɯ] 'strip of cloth' ⇨ [kíl̥**$ú_F$** kíl̥**á**ɯ] 'strips of cloth, etc.'

- Fixed CV in Tamil [$ki(ː)_F$] (Keane 2006)
  - [paɳam] 'money' ⇨ [p**a**ɳam **ki_F**ɳam] 'money, etc.'
  - [maːʈu] 'cattle' ⇨ [m**a**ːʈu **$kiː_F$**ʈu] 'cattle, etc.'

ECHO REDUPLICATION

- Unlike prototypical reduplication, echo reduplication typically **requires** the base and RED to be **non-identical**
  - Unlike "emergence of the unmarked" cases of base-RED nonidentity, e.g. Sanskrit (Steriade 1988)
  - Unlike "default fixed segmentism", e.g. Yoruba (Alderete et al. 1999)

- Presence of the fixed segment should be enough to generate base-RED nonidentity...
- ...unless the fixed segment is identical to the segment it is meant to replace

IDENTITY AVOIDANCE

- [m]-initial words in Turkish [m_F] have **no echo form**
  - [pɑrɑ] 'money' ⇨ [pɑrɑ m_Fɑrɑ] 'money, etc.'
  - [mɑsɑ] 'table' ⇨ *[mɑsɑ m_Fɑsɑ] 'towel, etc.' ⇨ **NO OUTPUT**

- [m]-initial words in Abkhaz [m_F] take **backup** [$tʃ_F$] (Vaux 1996)
  - [gádʒak'] 'fool' ⇨ [gádʒak' $m_F$ádʒak'] 'fool, etc.'
  - [maát] 'money' ⇨ *[maát $m_F$aát] ⇨ [maát $tʃ_F$aát] 'money, etc.'

- In Classical Tibetan [a_F], **base takes backup** [o_F] (Beyer 1992)
  - [ndzog] ⇨ [ndza_Fg ndzog] 'jumbled up'
  - [glen] ⇨ [gla_Fn glen] 'very stupid'
  - [ŋan] ⇨ *[ŋa_Fn ŋan] ⇨ [ŋan ŋo_Fn] 'miserable'

IDENTITY AVOIDANCE

- Through various means, lgs work to **avoid categorical identity** between base and RED in echo forms
- Survey of echo forms in >100 lgs of India found **identity avoidance** in every case (Trivedi 1990)

- Previous work on echo forms generally describe a straightforward case of **categorical identity avoidance**
- No one has yet confirmed that this avoidance pattern does not extend to natural classes, or that it is not gradient

A PUZZLE FROM ENGLISH

- What about English [ʃm$_F$]?
  - [dɑktɚ] 'doctor' ⇨ [dɑktɚ ʃm$_F$ɑktɚ] 'doctor$_{\text{DISMISSIVE}}$'
  - [skul] 'school' ⇨ [skul ʃm$_F$ul] 'school$_{\text{DISMISSIVE}}$'
- Online survey, 190 respondents (Nevins & Vaux 2003)
- **Identity avoidance**: 95–97% of speakers rejected echo forms with [ʃm$_F$] for the 3 [ʃm]-initial words
  - [ʃmuz] 'schmooze' ⇨ *[ʃmuz ʃm$_F$uz] 'schmooze$_{\text{DISMISSIVE}}$'
- Interestingly, 30% of speakers also rejected echo forms with [ʃm$_F$] for the one [ʃn]-initial word… **why??**
  - [ʃnɑz] 'schnozz' ⇨ *[ʃnɑz ʃm$_F$ɑz] 'schnozz$_{\text{DISMISSIVE}}$'

A PUZZLE FROM ENGLISH

- Possible explanations:

- The "two dialects" possibility
  - 65% of subjects obey **identity avoidance**
  - 30% obey **categorical similarity avoidance**, where [ʃn] and [ʃm] are of the same category: "sounds similar to [ʃm$_F$]"

- The "matter of degree" possibility
  - 95% obey **gradient similarity avoidance**, of whom:
  - 65% considered [ʃn] and [ʃm$_F$] are sufficiently dissimilar
  - 30% considered [ʃn] and [ʃm$_F$] are excessively similar

A PUZZLE FROM ENGLISH

- Another possible explanation: "this isn't English"
  - **Humorous** and possibly **peripheral** to the language
  - **Less common** in English than in other lgs
  - [ʃm] is **highly marked**, restricted to **borrowings** from Yiddish
  - Construction is possibly borrowed from Yiddish (Southern 2005)

MOTIVATION

- To understand if echo reduplication can employ gradient similarity avoidance, we need a lg in which:
  - Echo reduplication is a fully **productive**, **linguistic** feature
  - The fixed segment is a relatively **unmarked sound**
  - The fixed segment has many **similar** sounds

- **Bengali[1] is an ideal test case**
  - Default fixed segment [t_F][2]: crosslinguistically unmarked
  - [t] has high token freq. (definite marker & classifier [-ta])
  - Attested backup fixed segments [m_F f_F p_F u_F] (Ray et al. 1966)
  - Inventory has many [t]-like sounds: [t^h ɖ $ɖ^ɦ$ ʈ $ʈ^h$ ɖ̤ tɕ s…] (Khan 2010)

[1] Specifically, urban colloquial Bangladeshi varieties
[2] [t t^h ɖ $ɖ^ɦ$] can be retroflex in Bengali, but are typically alveolar in these varieties (Khan 2010)

QUESTIONS

- Does echo reduplication in Bengali involve…
  - **Categorical identity avoidance**,
  - **Categorical similarity avoidance**, or
  - **Gradient similarity avoidance**?
- If it is the latter, how can similarity be objectively **measured** on a gradient scale?
- As a comparison, we can investigate other parts of Bengali phonology that expected to employ this gradient similarity:
  - Lexical **cooccurrence** restrictions
  - Perceptual **confusability**

EXPERIMENT I: PRODUCTION

- **Basic design:** native speakers produce echo RED for base stimuli with carefully-selected initial C

- **Expectations:**
  - [kaʃi] 'cough' ⇨ [kaʃi t_Faʃi]
  - [$b^ɦ$idʑːa] 'having gotten wet' ⇨ [$b^ɦ$idʑːa $t_F$idʑːa]
  - [tika] 'vaccine' ⇨ *[tika $t_F$ika] ⇨ [tika $m_F$ika] (identity violation)

- **Question: how will sounds similar to [t_F] behave?**
  - [t^hajʃːa] 'having stuffed' ⇨ [t^hajʃːa $t_F$ajʃːa] (no violation)?
    OR
  - [t^hajʃːa] ⇨ *[t^hajʃːa $t_F$ajʃːa] ⇨ [t^hajʃːa $f_F$ajʃːa] (similarity violation)?

EXPERIMENT I: STIMULI

- **60 stimulus** words
  - Disyllabic stems
  - Content words: N, A, V (perfective participles)
- **2 registers** of urban colloquial Bangladeshi Bengali
  - High register: closer to written Kolkata Standard
  - Low register: closer to eastern regional varieties
- **Produced** by adult female speaker
  - Proficient in both registers
  - 2 reps per variety = 240 recordings
  - Recorded in sound-treated booth on Telex M-540 mic

EXPERIMENT I: STIMULI

- 60 test words fell under **3 conditions**:
- **Identity**: [t]-initial words
- **Similarity**: words with [t]-like initials
  - Coronal obstruents [tʰ d t̪ t̪ʰ tɕ s~tɕʰ ʃ]
- **Control**: words with non-[t]-like initials
  - Coronal sonorants [n l ɹ]
  - Non-coronals [k h p f bʱ m]

EXPERIMENT I: STIMULI

- Consonants of Bangladeshi Standard Bengali (Khan 2010)

Identity **Similarity** **Control**

	Labial	Dental	Alveolar	Post-Alv	Velar/Glot
Stop	p b $b^{ɦ}$	t̪ $t̪^{h}$ d̪ $d̪^{ɦ}$	t t^{h} d $d^{ɦ}$		k k^{h} g $g^{ɦ}$
Affricate			tɕ $tɕ^{h}$ dʑ $dʑ^{ɦ}$		
Fricative	f	s		ʃ	h
Liquid		l ɹ			
Nasal	m	n			(ŋ)

EXPERIMENT I: STIMULI

- Consonants of Bangladeshi Standard Bengali (Khan 2010)

Identity **Similarity** **Control**

	Labial	Dental	Alveolar	Post-Alv	Velar/Glot
Stop	p b $b^{ɦ}$	t̪ $t̪^{h}$ d̪ $d̪^{ɦ}$	t t^{h} d $d^{ɦ}$		k k^{h} g $g^{ɦ}$
Affricate			tɕ $tɕ^{h}$ dʑ $dʑ^{ɦ}$		
Fricative	f		s	ʃ	h
Liquid			l ɹ		
Nasal	m		n		(ŋ)

EXPERIMENT I: SETUP

- **30 speakers** of Bengali
  - Varied dialect background
  - Residents of CA
  - Paid $10

- **Heard stimulus**
  - Participant selected preferred register
  - Order randomized for each speaker
- **Asked to produce** echo reduplicated form
  - [kaʃi] 'cough' ⇨ [kaʃi t_Faʃi] 'cough, etc.' given as example
- Responses were transcribed

EXPERIMENT I: HYPOTHESES

- **Identity** words will never use [t_F]
- **Control** words will always use [t_F]
- **Similarity** words are what are being tested:
  - Hypothesis 1: **similarity** = **control** (categorical identity)
  - Hypothesis 2: **similarity** = **identity** (categorical similarity)
  - Hypothesis 3: **similarity** is on a continuum

Identity		**Similarity**		**Control**
*[t…t_F]	≠	[t^h…t_F]	=	[$b^ɦ$…t_F]

[t^hajʃːa] ⇨ [t^hajʃːa $t_F$ajʃːa]

EXPERIMENT I: HYPOTHESES

- **Identity** words will never use [t_F]
- **Control** words will always use [t_F]
- **Similarity** words are what are being tested:
  - Hypothesis 1: **similarity** = **control** (categorical identity)
  - Hypothesis 2: **similarity** = **identity** (categorical similarity)
  - Hypothesis 3: **similarity** is on a continuum

Identity		**Similarity**		**Control**
*[t…t_F]	=	*[t^h…t_F]	≠	[$b^ɦ$…t_F]

[t^hajʃːa] ⇨ *[t^hajʃːa $t_F$ajʃːa] ⇨ [t^hajʃːa $m_F$ajʃːa]

EXPERIMENT I: HYPOTHESES

- **Identity** words will never use [t_F]
- **Control** words will always use [t_F]
- **Similarity** words are what are being tested:
  - Hypothesis 1: **similarity** = **control** (categorical identity)
  - Hypothesis 2: **similarity** = **identity** (categorical similarity)
  - Hypothesis 3: **similarity** is on a continuum

Identity		**Similarity**		**Control**
*[t…t_F]	≠	?[t^h…t_F]	≠	[$b^ɦ$…t_F]

[t^hajʃːa] ⇨ [t^hajʃːa $t_F$ajʃːa] ~ [t^hajʃːa $f_F$ajʃːa]

EXPERIMENT I: RESULTS

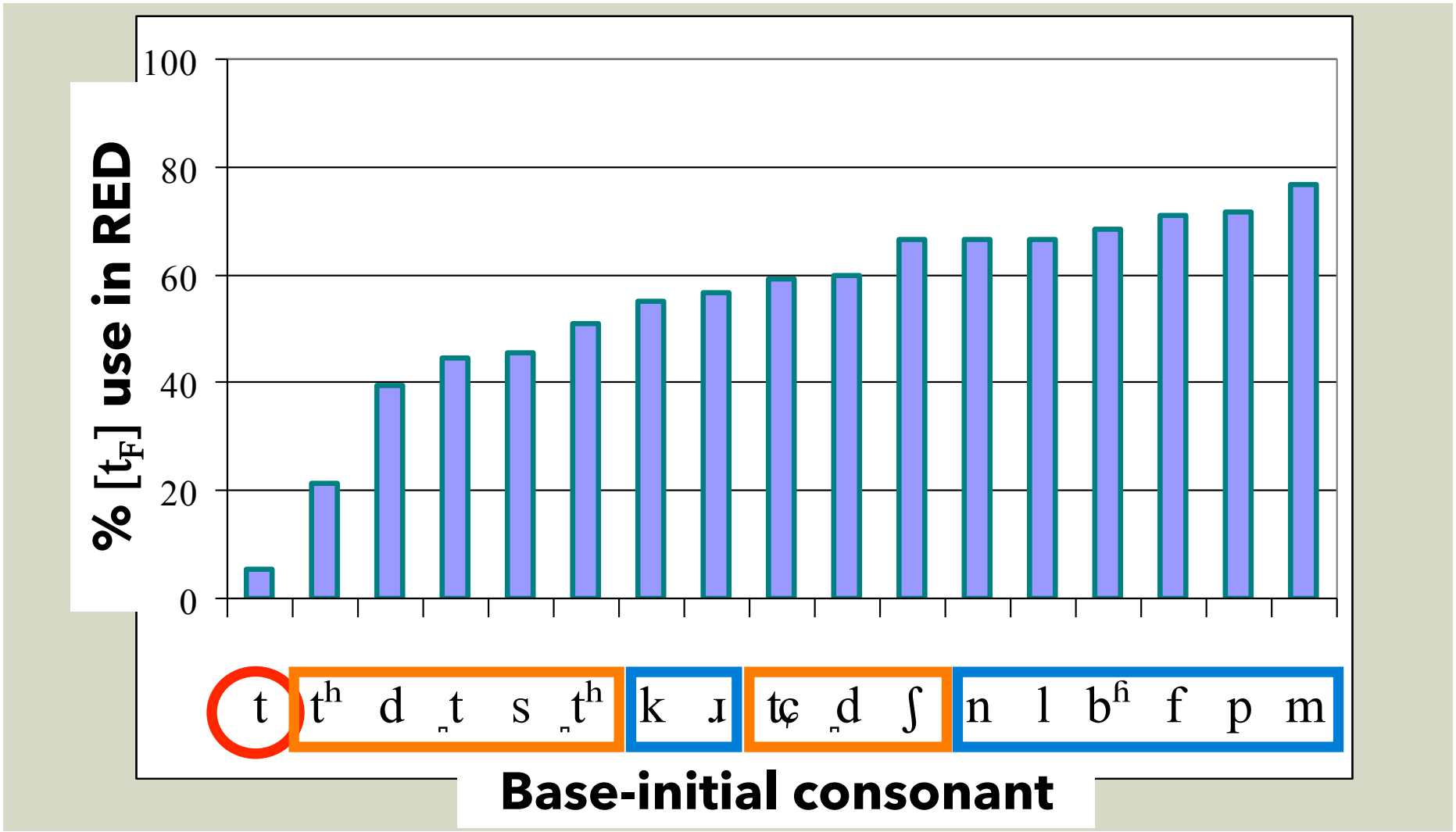
- **Hypothesis 3** was borne out
- Similarity words lie on a continuum
  - Disprefer [t_F] but not outright ungrammatical
  - Some consonants are more [t]-like in behavior than others
- Seems like **Cs that take [t_F] less often are also phonetically closer to** [t]

t t^h d $\underset{\sqcap}{t}$ ... b^{fi} f p m

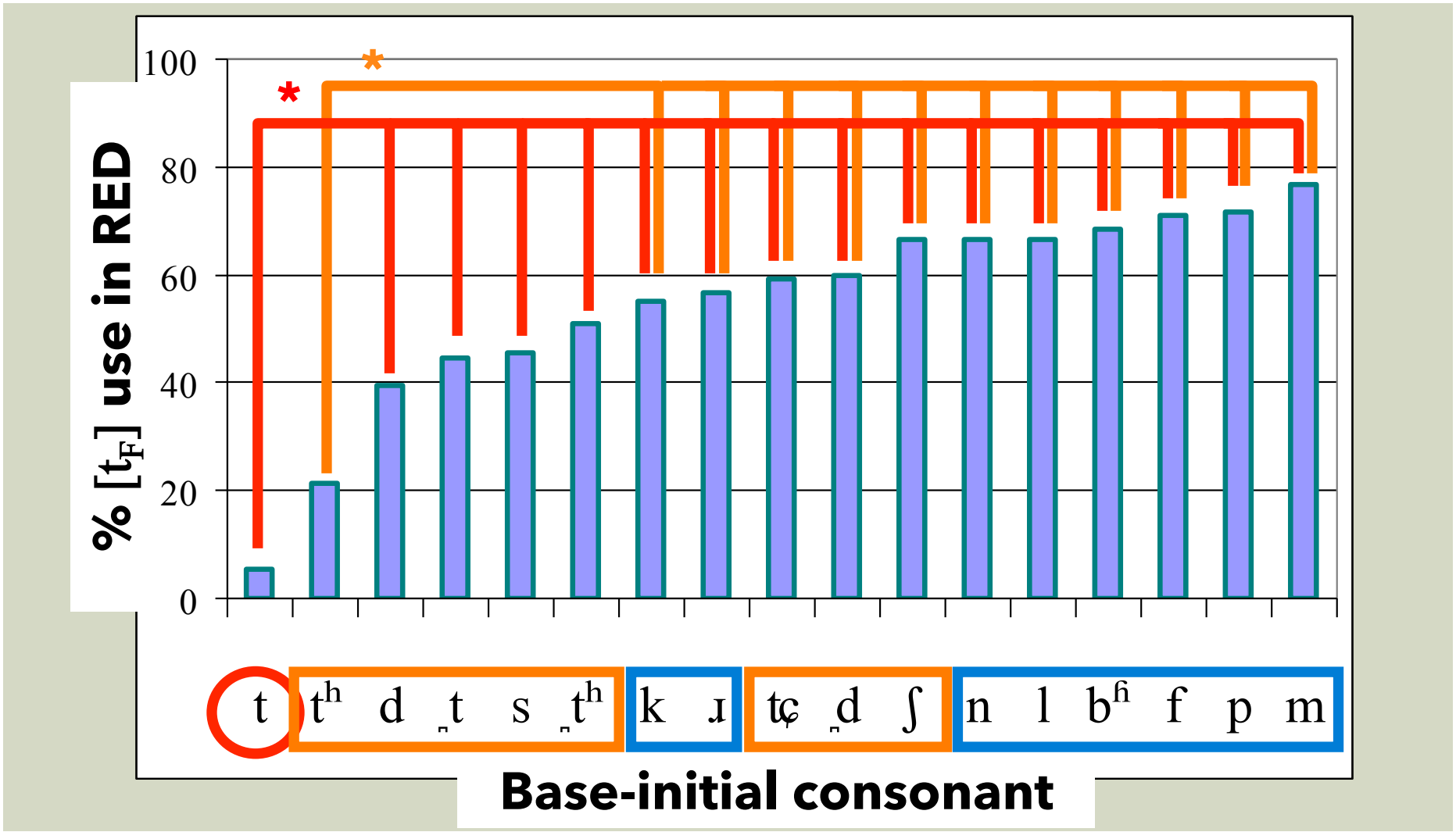
Least likely to be replaced by [t_F]

Most likely to be replaced by [t_F]

EXPERIMENT I: RESULTS



EXPERIMENT I: RESULTS



EXPERIMENT I: DISCUSSION

- Echo reduplication in Bengali appears to incorporate a notion of **gradient similarity avoidance**
  - No straightforward clustering of consonants
  - Heavy overlap across clusters
  - Like the "matter of degree" hypothesis from English puzzle

NEW QUESTIONS

- We should confirm our suspicion that our reduplication data can be modeled on an **objective scale of similarity**
- Is there a **metric** that Bengali speakers are using to calculate the similarity of an initial C and [t]?
- Metric has to be **gradient**, possibly **language-specific**

SHARED NATURAL CLASSES

- Best-known option is **shared natural classes (SNC) metric** (Frisch et al. 1995/2004)
- **Similarity** of two Cs is based on the **number of natural classes they share** in the inventory
- Universal claim with language-specific application

- **Hypothesis:** the more natural classes shared between a C and [t], the less likely it will take [t_F] in its echo RED

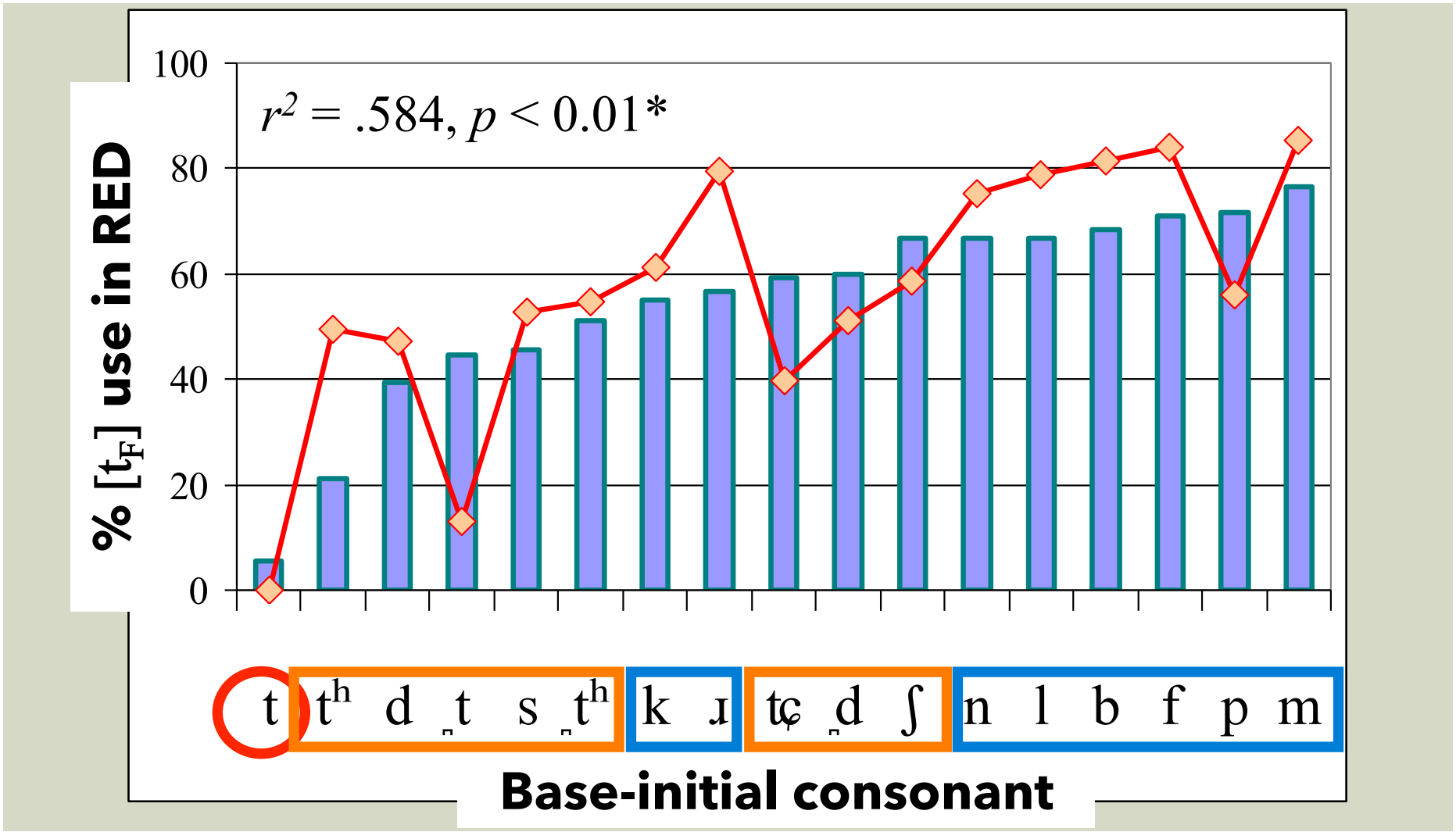
SNC: METRIC

- In the SNC metric, similarity of C_1 and [t] is quantified as:

$$sim(C_1, t) = \frac{\#\text{ natural classes shared by } (C_1, t)}{\#\text{ shared natural classes} + \#\text{ non-shared natural classes}}$$

- Compared SNC-similarity (line) to Exp 1 results (bars)

SNC: CORRELATION



SNC: DISCUSSION

- The SNC metric does an okay job overall ($r^2 = .584$)
- However, the area where it crucially **fails** to predict the data is the **similarity set** (coronal obstruents)

- The metric treats [t] as inherently more similar to [t̪] and [tɕ] than to [t^h]... is there a way to adjust that?

SNC: THOUGHT EXPERIMENT

- Original SNC metric derives directly from the phoneme inventory and feature set
- But what if we maintain the basic model but incorporate **feature weights**?

- Let's try a little **thought experiment**
- Weighting [dist] over [spread gl]: the [t - t̪] distinction can be "heavier" than the [t - t^h] distinction
- If this improves our metric, we can then pursue the question of whether these weights are justified

WEIGHTED SNC: METRIC

- In an SNC-like model with feature **weights**, similarity of C_1 and [t] is quantified as follows: (Wilson, p.c.)

$$sim(C_1, t) = \exp\left(-\sum_{i=1}^{\#features} w_i(1-\delta_i(C_1, t))\right)$$

w_i = weight of the feature f_i

$\delta_i(C_1, t) = 1$ (feature value shared) or 0 (not shared)

- Where weights are drawn from the variation in the reduplication results, as follows:

WEIGHTED SNC: METRIC

- Probability of [t_F] use in the RED of a base with initial C_1

$$P = ((m!) \div (n!(m-n)!)\ (1-sim(C_1, t))^n\ (sim(C_1, t))^{m-n}$$

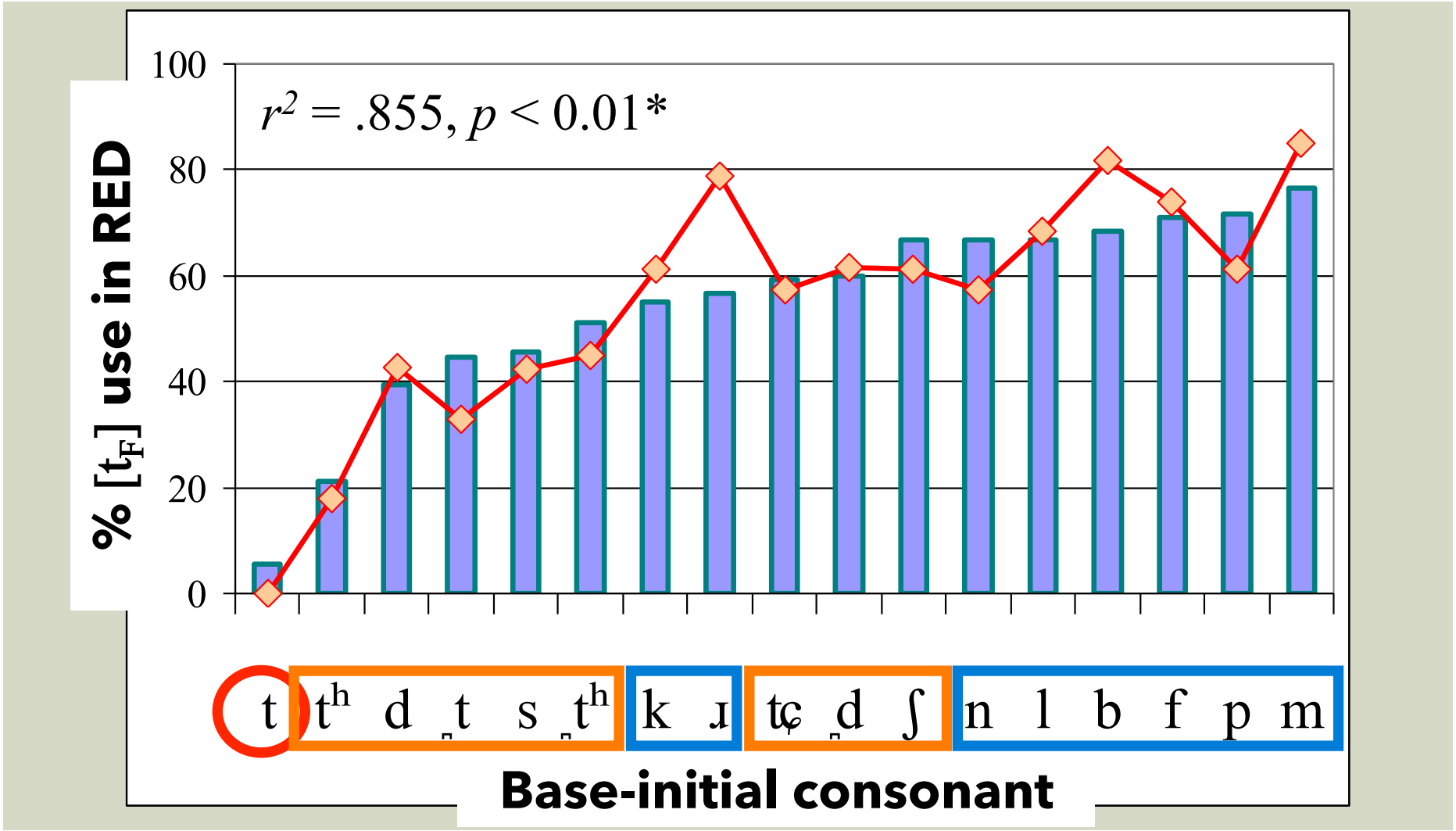
P = probability that C_1-initial base will be reduplicated with [t_F] *n* times out of a total of *m* trials

m = number of reduplications for C_1-initial word

n = number of reduplications with [t_F] for C_1-initial word

- Compared weighted similarity (line) to Exp 1 results (bars)

WEIGHTED SNC: CORRELATION



WEIGHTED SNC: DISCUSSION

- With **4 adjusted feature weights**, the SNC metric can closely model the reduplicative data ($r^2 = .855$)
  - [voice]: .554
  - [distributed]: .400
  - [strident]: .249
  - [spread glottis]: .198
  - All other features have a weight of 0.100

NEW QUESTION

- Okay, but have we compromised the model?
- Is it no longer a similarity metric, but just a model of the reduplicative data?

- Let's see if our reduplicative data resemble other areas where gradient, lg-specific similarity is arguably relevant:
  - Lexical **cooccurrence** (McCarthy 1994)
  - Perceptual **confusability** (Shepard 1972)

COOCCURRENCE

- **Similarity** of two Cs is often **negatively correlated** with their **cooccurrence** within roots (Greenberg 1950)
  - English: two LAB or two DOR are underattested in [sCVC]: skip, speak, skim, smack…, *smap, *scog, *spobe, *speam (Fudge 1969)
  - Arabic: velars & uvulars rarely cooccur within roots (Frisch et al. 2004)

- **Hypothesis**: the less often a C cooccurs with [t] in a root, the less often it will take [t_F] in its echo RED
- If we see a strong correlation with the reduplicative data, this could be independent support for our weighted model

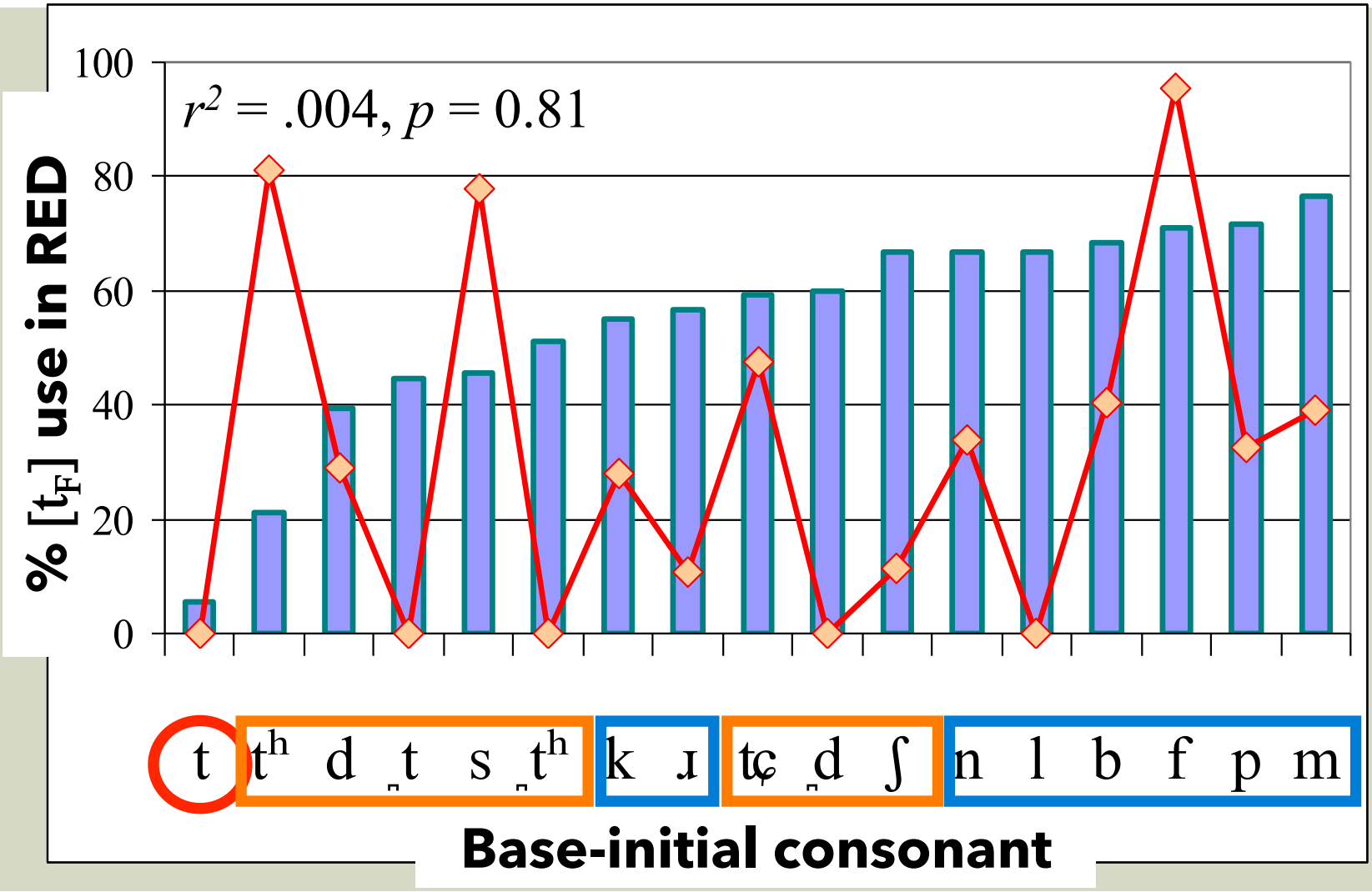
COOCCURRENCE: METRIC

- Similarity of initial C_1 and medial [t] is the inverse of their observed / expected lexical cooccurrence: (Frisch et al. 2004)

$$sim(C_1, t) = \frac{\frac{\#[C_1VCV]}{\#[CVCV]} \times \frac{\#[CVtV]}{\#[CVCV]}}{\frac{\#[C_1VtV]}{\#[CVCV]}}$$

- Examined the cooccurrence of all initial Cs with medial [t] in CVCV roots in a corpus of Bengali (Mallik et al. 1998)
- Compared cooccurrence rate (line) to Exp 1 results (bars)

COOCCURRENCE: CORRELATION



COOCCURRENCE: DISCUSSION

- The lexical cooccurrence model of similarity **fails to predict** the observed [t_F]-avoidance patterns ($r^2 = .004$)

- Possible explanations:
- Lexical cooccurrence in Bengali involves similarity, but echo reduplication does not (unlikely, see results)
- Lexical cooccurrence in Bengali does not involve similarity, while echo reduplication does (possible)
- **Low *n*?** Corpus had 865 CVCV roots; 64 with medial [t]
  - cf. Arabic corpus of 2674 roots (Frisch et al. 2004)

CONFUSABILITY

- The other area to look for the effects of gradient similarity is in **perceptual confusability**
  - Hindi: [ʈ] is misidentified as [ʈ̪] more than as [ɖ] (Ahmed & Agrawal 1968)

- **Hypothesis**: Cs more likely to be (mis)perceived as [t] are also less likely to take [t_F] in echo RED
- If we see a strong correlation with the reduplicative data, this could be independent support for our weighted model

EXPERIMENT II: SETUP

- **Multiple Forced Choice** (MFC) listening experiment
  - Participants identify the consonant they hear
  - Run in Praat (Boersma & Weenink 2013)
  - Sony MDR-V200 headphones connected to laptop
  - Experiments took place in quiet room in participants' homes

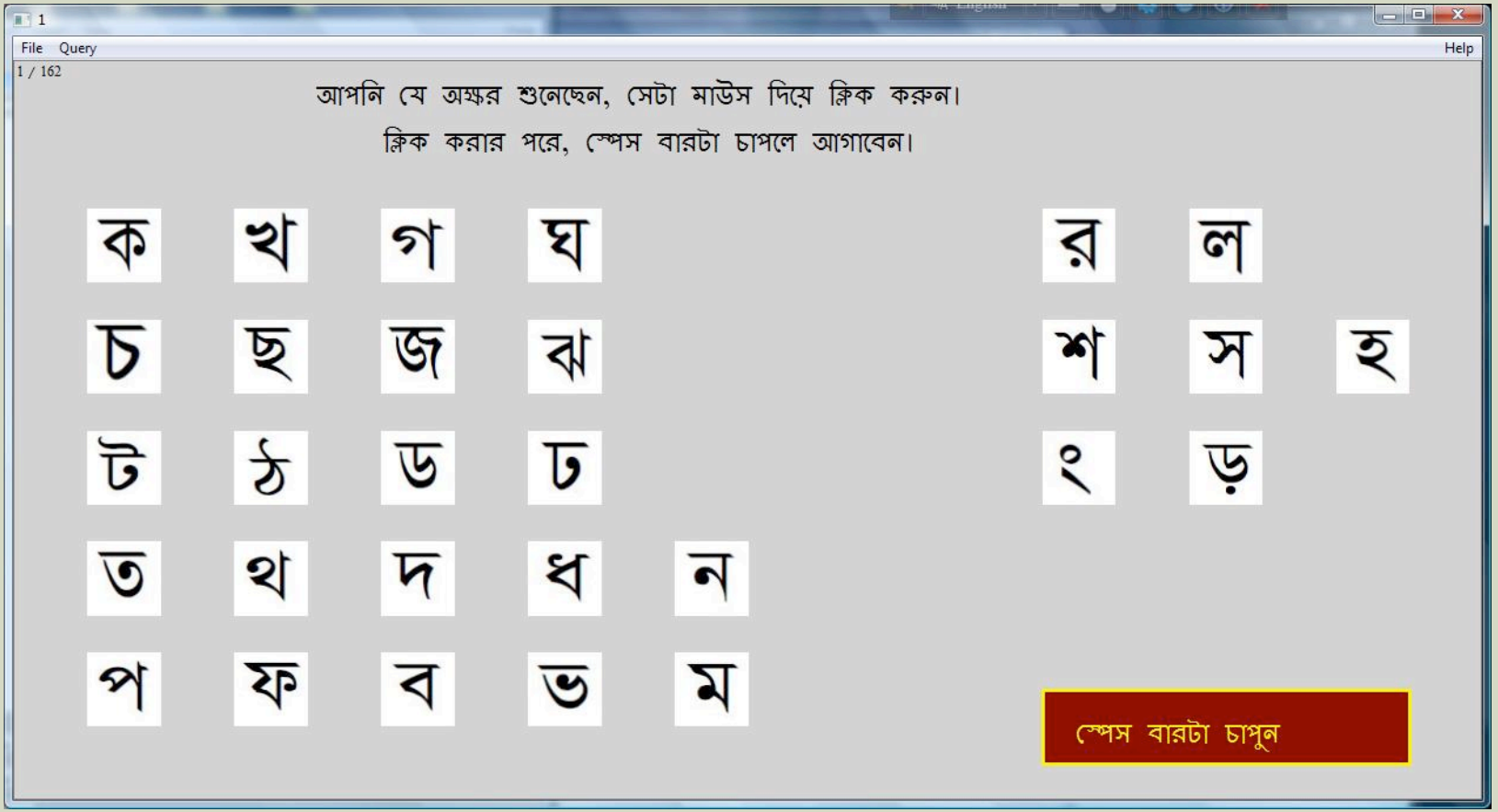
- **25 speakers** of Bengali (13F, 12M)
  - Reported no hearing difficulties
  - Varied dialect background
  - Residents of or visitors to CA
  - Paid $20

EXPERIMENT II: STIMULI

- **54 syllables**
  - Onsets: 27 legal [Ca] syllables (all Cs but [ŋ ɽ])
  - Codas: 27 legal [aC] syllables (all Cs but [dˤ ɦ])
- Produced by adult female speaker
  - Best of several reps was normalized for amplitude

- **Blocked** by **3 masking conditions**
  - Multi-talker **babble**
  - Pink **noise**
  - **Quiet** (no added sound)
- 54 syllables x 3 conditions x 3 reps = **486 trials**

EXPERIMENT II: TASK

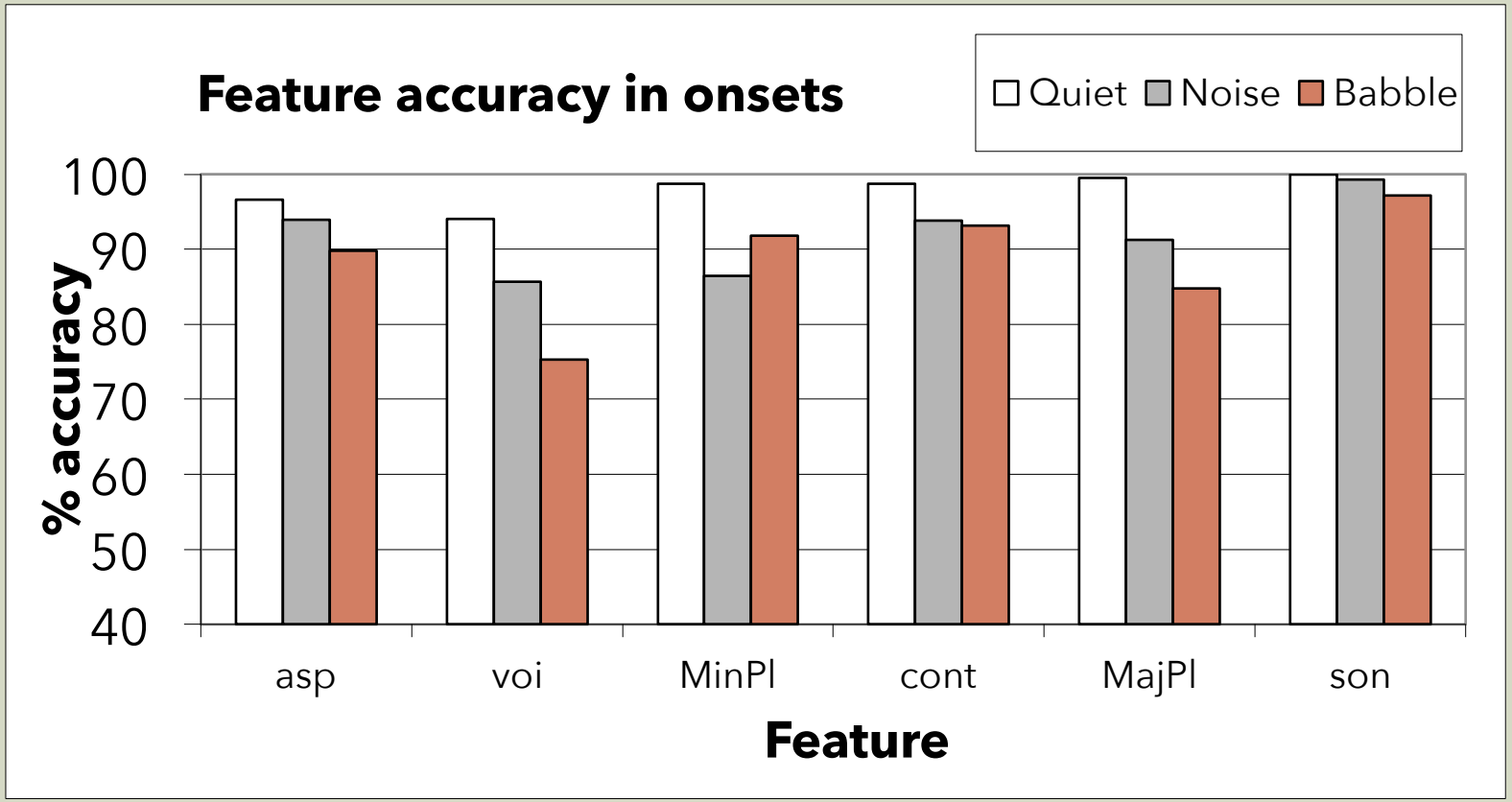


EXPERIMENT II: HYPOTHESES

- The C most confused with [t] should be [t^h]
  - Generalized: **aspiration** should be the **most confusable** feature
- Next most confused with [t] should be [d]
  - Generalized: **voicing** should be the **2nd most confusable** feature
- After that should be [t̪]
  - Generalized: **[distributed]** and other **minor place** distinctions should be the **3rd most confusable**
- After that should be [s]
  - Generalized: **[strident]** and other **manner-related** distinctions should be **less confusable** than the preceding

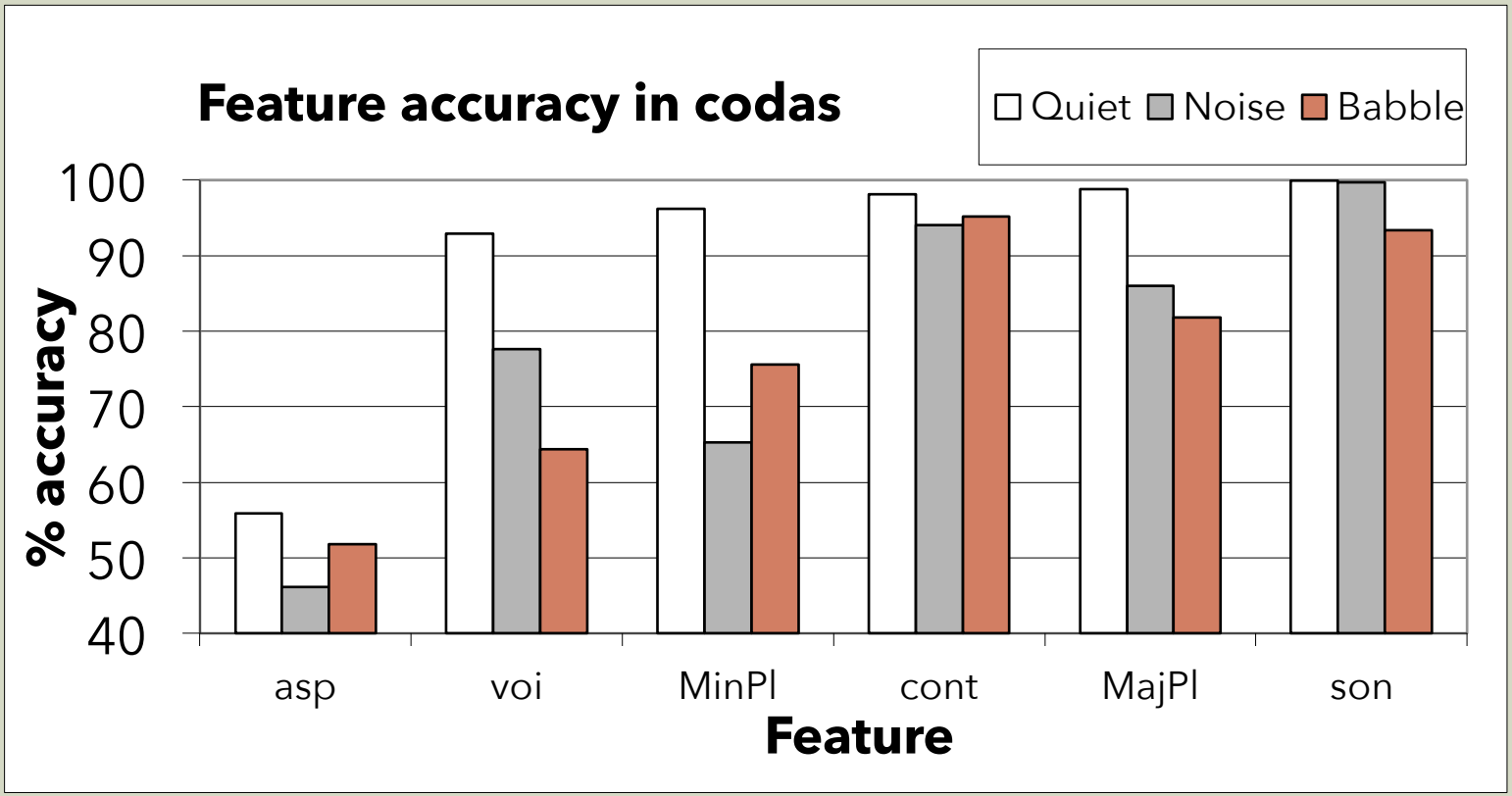
EXPERIMENT II: RESULTS

- **Onset accuracy:** 92% in quiet, 70% in noise, 60% in babble



EXPERIMENT II: RESULTS

- **Coda accuracy**: 66% in quiet, 39% in noise, 34% in babble



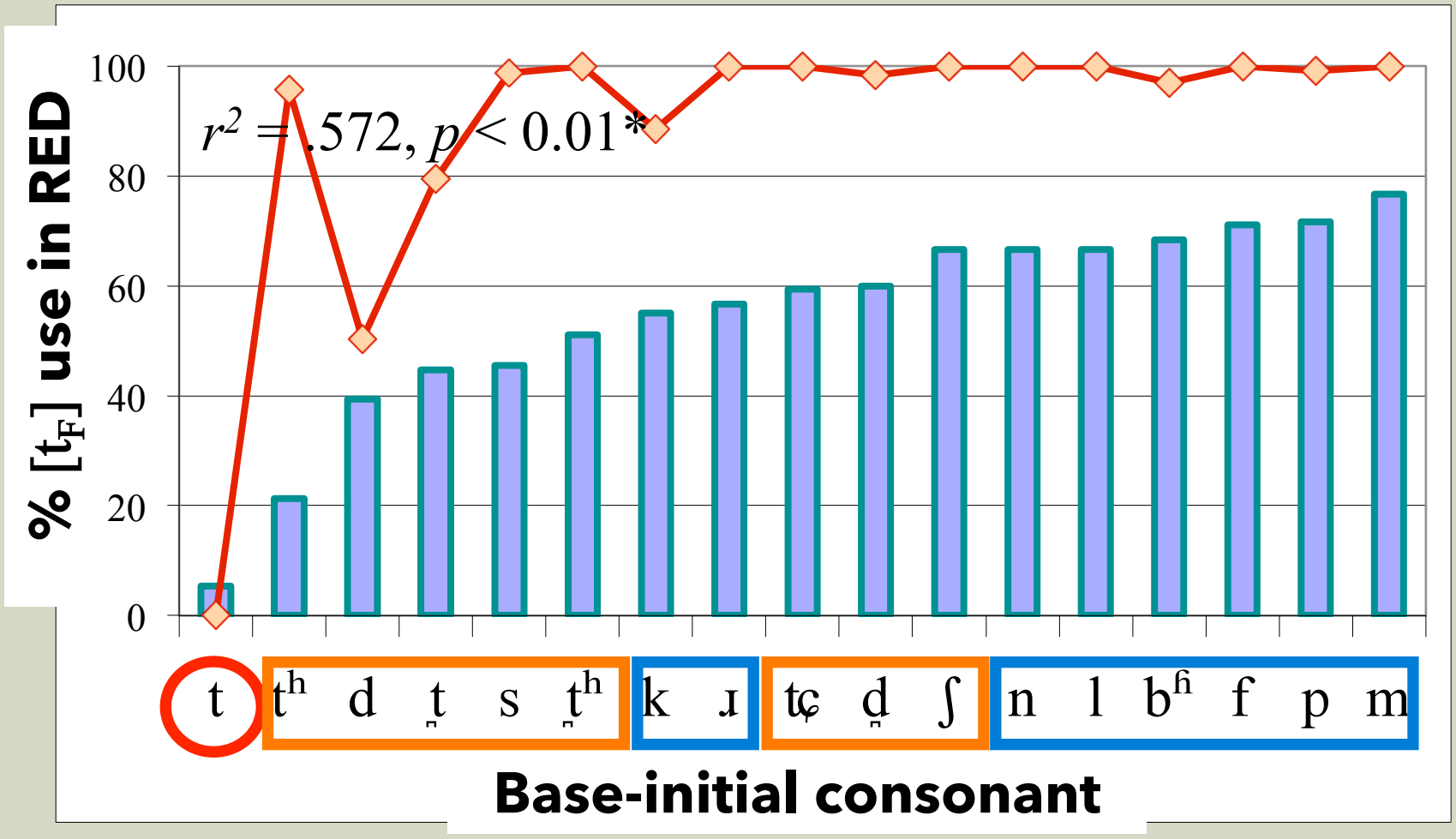
CONFUSABILITY: METRIC

- Similarity of C_1 and [t] as drawn from confusion rate is quantified as follows: (Shepard 1972)

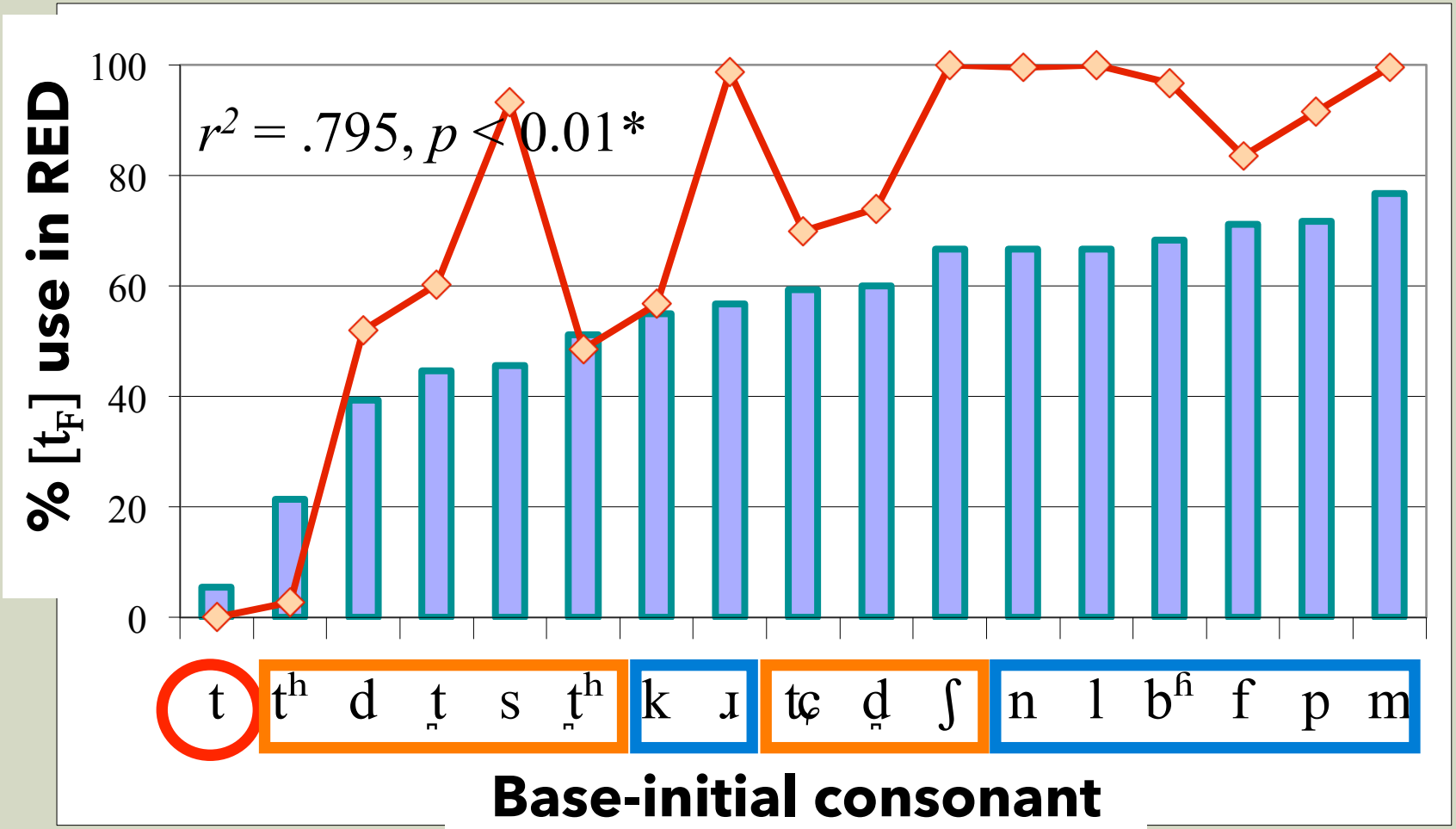
$$sim(C_1, t) = \frac{\#(C_1{:}t) + \#(t{:}C_1)}{\#(C_1{:}C_1) + \#(t{:}t)}$$

- Compared Exp 2 perceptions to Exp 1 productions
  - Removed "quiet" condition results (at ceiling)
  - Looked at onsets and codas separately

ONSET CONFUSIONS: CORRELATION



CODA CONFUSIONS: CORRELATION



CONFUSABILITY: DISCUSSION

- Consonant **confusions with** [t] **in coda position** are **well correlated** with the reduplicative results ($r^2 = .795$)

- But! Echo reduplication involves judging the **similarity of onsets**; why does the reduplicative data more closely resemble **coda confusion**?
  - Onset confusions with [t] were overall rare
  - Acoustic cues are perceptually less salient in codas (Wright 2004), so this is where similarity (not just identity) is likely more often relevant

SYNTHESIS OF RESULTS

- Okay, we need a recap.

- What did we do again?
  - Task 1: examine **fixed segment choice** in echo reduplication
  - Task 2: establish that fixed segment choice is **predicted by SNC**
  - Task 3: improve the SNC in a thought experiment with **weights**
  - Task 4: find no correlation with **lexical statistics**
  - Task 5: find significant correlation with **coda confusions**

CONCLUSIONS

- The current study demonstrates that fixed segment choice in Bengali echo reduplication is **highly variable**
- I argue that the choice of fixed segment involves a systematic avoidance of **similarity**, because:
  - The patterns are (partially) predicted by the **SNC metric**
  - The patterns correlate with **confusion rates** (in codas)

CONCLUSIONS

- The patterns clearly show that this similarity is **gradient**
- Many other phenomena previously treated as categorical have since been shown to be gradient
  - e.g. vowel harmony in Hungarian (Hayes & Londe 2006)

CONCLUSIONS

- The current study proposes a modified version of the SNC metric of similarity
- **I propose feature weighting** for lg-specific application in diverse phonological phenomena

- The study also provides an interesting case in which the SNC metric can measure similarity in phonological phenomena *other than* lexical cooccurrence effects

REMAINING QUESTIONS

- Is Bengali echo reduplication a special case, or should we look for gradient similarity in many more lgs?
- Why are the lexical cooccurrence effects of Bengali so different from the reduplicative results?
- How does this change as speakers deal with multiple phoneme inventories, e.g. bilinguals?

ACKNOWLEDGMENTS

- This study was supported in part by Reed College's Stillman Drake Fund.
- To my participants and stimulus producers, to Colin Wilson (JHU), Kie Ross Zuraw (UCLA), Marc Garellek (UCSD), and Megha Sundara (UCLA), and to the audience at IIT Delhi:

অসংখ্য ধন্যবাদ!

[ɔʃoŋkʰo ɖʱonːobaɖ]

REFERENCES

Ahmed, Rais; Agrawal, S. S. 1968. Significant features in the perception of (Hindi) consonants. *JASA* 45(3).

Alderete, John; Beckman, Jill; Benua, Laura; Gnanadesikan, Amalia; McCarthy, John; Urbanczyk, Suzanne. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30(3), 327-364.

Beeler, Madison S. 1970. Sibilant harmony in Chumash. *IJAL* 36(1), 14-17.

Beyer, Stephan V. 1992. *The Classical Tibetan language*. SUNY Press.

Boersma, Paul; Weenink, David. 2013. Praat: doing phonetics by computer, version 5.3.39, http://www.praat.org/

Coetzee, Andries; Pater, Joe. 2005. Gradient phonotactics in Muna and Optimality Theory. Ms.

Cutler, Ann; Weber, Andrea; Smits, Roel; Cooper, Nicole. 2004. Patterns of English phoneme confusions by native and non-native listeners. *JASA* 116(6), 3668-3678.

Frisch, Stefan; Pierrehumbert, Janet; Broe, Michael. 1995/2004. Similarity avoidance and the OCP. *NLLT* 22, 179-228.

Fudge, E. C. 1969. Syllables. *Journal of Linguistics* 5, 253-287.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6, 162-181.

Hayes, Bruce; Londe, Zsuzsa Cziráky. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23(1), 59-104.

Khan, Sameer ud Dowla Khan. 2010. Bengali (Bangladeshi standard). *JIPA* 40(2), 221-225.

Mallik, Bhakti Prasad; Bhattacharya, Nikhilesh; Kundu, Subhas Chandra; Dawn, Mina. 1998. *The phonemic and morphemic frequencies of the Bengali language*. Calcutta: The Asiatic Society.

REFERENCES (CONT.)

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Keating, Patricia (ed.) *Phonological structure and phonetic form: papers in laboratory phonology 3*, 191-233.

Mortensen, David. 2006. Logical and substantive scales in phonology. UC Berkeley PhD dissertation.

Nevins, Andrew; Vaux, Bert. 2003. Metalinguistic, shmetalinguistic: the phonology of shm-reduplication. Ms.

Ray, Punya Sloka; Hai, Muhammad Abdul; Ray, Lila. 1966. *Bengali language handbook*. Washington: Center for Applied Linguistics.

Shepard, Roger N. 1972. Psychological representation of speech sounds. In David, E. E.; Denes, P. B. (eds.) *Human communication: a unified view*, 67-113.

Southern, Mark. 2005. *Contagious couplings: transmission of expressives in Yiddish echo phrases.* Westport: Greenwood.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5(1), 73-155.

Tosun, Gülsat. 1999. Vowel harmony in Turkish and Turkmen. Ms.

Trivedi, G. M. 1990. Echo Formation. In Krisha, Shree (ed.) *Linguistic Traits Across Language Boundaries*. Anthropological Survey of India. Calcutta: Ministry of Human Resource Development.

Vaux, Bert. 1996. Abkhaz Mabkhaz: m-reduplication in Abkhaz and the problem of melodic invariance. Ms.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Hayes, Bruce; Kirchner, Robert; Steriade, Donca (eds.) *Phonetically Based Phonology*. Cambridge.