Imperfect Rhymes as a Measure of Phonological Similarity

Aaron Kaplan & John Woodmansee
a.kaplan@utah.edu

AK: Univ. of Utah; JW: NCSSM

April 7, 2018
UNC Spring Colloquium


NCSSM
NCSSM

Similarity

- The degree of similarity between segments is central to many domains of phonology:
  - IO & BR Faithfulness (McCarthy & Prince 1995)
  - OCP (e.g. Leben 1973): adjacent elements must be dissimilar
  - Agreement by Correspondence (e.g. Rose & Walker 2004): harmony between segments that meet a threshold of similarity
- Intuition: speakers are aware of and can measure how similar segments are. Sometimes similarity is avoided (OCP), and sometimes it is reinforced (ABC).

Measuring Similarity

- Similarity is measured with distinctive features, and all features are equal.
- Does this match speakers' intuitions about similarity?

Measuring Similarity

- Do more featural differences = greater dissimilarity?
- Is a difference in [±F] equivalent to a difference in [±G]?

Probing speakers' assessments about similarity: imperfect rhymes

Imperfect Rhymes

- Imperfect rhymes: sometimes rhyming words don't rhyme exactly:

This version of the world will not be here long [laŋ]
It is already gone It is already gone [gan]

T Bone Burnett, "Palestine, Texas"

- Assuming lyricists are more likely to use similar-sounding imperfect rhymes than dissimilar ones, we can use imperfect rhymes to probe speakers' judgments about segmental similarity.

Imperfect Rhymes

- If featural similarity matches speakers' judgments about similarity, the frequency of consonantal pairings in imperfect rhymes should be inversely proportional to the number of features they mismatch on.

The Data

- Zwicky (1976): a limited study of “rock rhyme” in 1960s–1970s rock.
- Our study: rhymes from 117 songs from many genres of popular music; 1977–2016.
- Data collected by AK and students at the North Carolina School of Science and Mathematics.
  - Juniors in John Woodmansee & Ormand Moore’s 2016–2017 American Studies class

The Data

For today, 294 rhyming pairs of words meeting the following criteria:

- “Masculine” rhymes: the stressed/rhyming syllables are final: *unfair/compare*

The Data

For today, 294 rhyming pairs of words meeting the following criteria:

- "Masculine" rhymes: the stressed/rhyming syllables are final: *unfair/compare*
    - "Feminine" rhymes (*treble/rebel*): stressed syllable and all following syllables "should" match. Not sure how to handle them yet...

The Data

For today, 294 rhyming pairs of words meeting the following criteria:

- “Masculine” rhymes: the stressed/rhyming syllables are final: *unfair/compare*
    - “Feminine” rhymes (*treble/rebel*): stressed syllable and all following syllables “should” match. Not sure how to handle them yet. . .
- Identical vowels (analysis here focuses on consonants)

The Data

For today, 294 rhyming pairs of words meeting the following criteria:

- "Masculine" rhymes: the stressed/rhyming syllables are final: *unfair/compare*
  - "Feminine" rhymes (*treble/rebel*): stressed syllable and all following syllables "should" match. Not sure how to handle them yet...
- Identical vowels (analysis here focuses on consonants)
- Same number of consonants: *long/gone* but not *fun/fund*

The Data

- Identical pairs included unless the pair is repeated in identical lines (e.g. it's in the chorus).

The Data

- Identical pairs included unless the pair is repeated in identical lines (e.g. it's in the chorus).
- Transcriptions pulled from CMU Pronouncing Dictionary

The Data

- Identical pairs included unless the pair is repeated in identical lines (e.g. it's in the chorus).
- Transcriptions pulled from CMU Pronouncing Dictionary
- In two words with shape $\ldots VC_1C_2 \ldots C_n$, we compared C_1 to C_1, C_2 to C_2, etc.

The Data

- Identical pairs included unless the pair is repeated in identical lines (e.g. it's in the chorus).
- Transcriptions pulled from CMU Pronouncing Dictionary
- In two words with shape $\ldots VC_1C_2\ldots C_n$, we compared C_1 to C_1, C_2 to C_2, etc.
  - This doesn't account for cases where Word 1's C_1 matches Word 2's C_2, but it's a good first approximation.

The Data

- Identical pairs included unless the pair is repeated in identical lines (e.g. it's in the chorus).
- Transcriptions pulled from CMU Pronouncing Dictionary
- In two words with shape $\ldots VC_1C_2 \ldots C_n$, we compared C_1 to C_1, C_2 to C_2, etc.
  - This doesn't account for cases where Word 1's C_1 matches Word 2's C_2, but it's a good first approximation.

Total: 378 pairs of mismatched consonants

Evaluating Featural Similarity

- Our feature system: an "average" of commonly accepted systems, perhaps most similar to Hayes (2009).

Evaluating Featural Similarity

- Our feature system: an "average" of commonly accepted systems, perhaps most similar to Hayes (2009).
- Uncontroversial features: [syll, son, approx, voc (= cons), lat, nas, cont, voi]

Evaluating Featural Similarity

- Our feature system: an "average" of commonly accepted systems, perhaps most similar to Hayes (2009).
- Uncontroversial features: [syll, son, approx, voc (= cons), lat, nas, cont, voi]
- [delayed release] to distinguish stops from affricates (fricatives are [–d.r.], contra Hayes)

Evaluating Featural Similarity

- Our feature system: an “average” of commonly accepted systems, perhaps most similar to Hayes (2009).
- Uncontroversial features: [syll, son, approx, voc (= cons), lat, nas, cont, voi]
- [delayed release] to distinguish stops from affricates (fricatives are [–d.r.], contra Hayes)
- Place features: to avoid inflation of featural differences, we used [lab, dental, cor, pal, dor] instead of [lab, cor, dor] with many dependent place features.

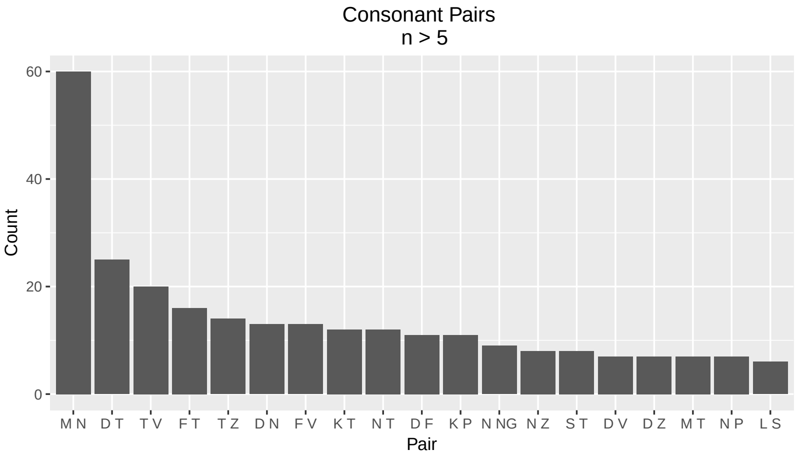
Evaluating Featural Similarity

- Our feature system: an "average" of commonly accepted systems, perhaps most similar to Hayes (2009).
- Uncontroversial features: [syll, son, approx, voc (= cons), lat, nas, cont, voi]
- [delayed release] to distinguish stops from affricates (fricatives are [–d.r.], contra Hayes)
- Place features: to avoid inflation of featural differences, we used [lab, dental, cor, pal, dor] instead of [lab, cor, dor] with many dependent place features.

⇒ This idealized feature system provides a rough starting point: do distinctive feature systems in general have a hope of reflecting speakers' judgments?

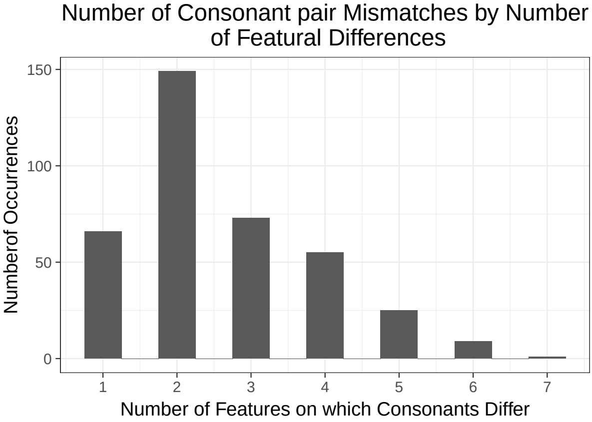
General Trends

Most Common Mismatched Consonants



Are mismatches with fewer featural differences more common?

Yes, mostly:

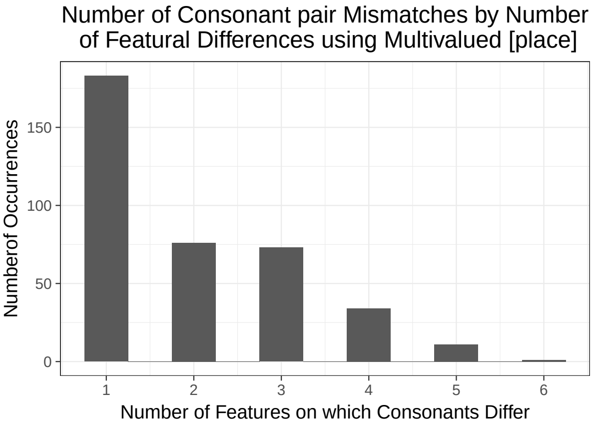


The Numbers & Some Examples

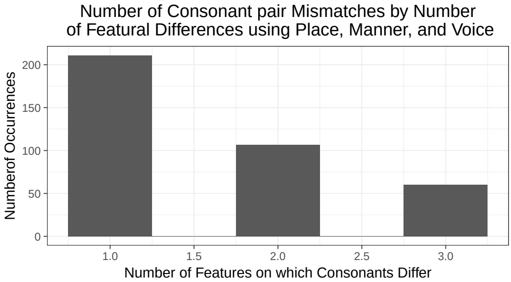
- One feature different: 66
- Two features different: 149
- Three features different: 73
- Four features different: 55
- Five features different: 25
- Six features different: 9
  - *smile/time* × 2; *while/time* (Colbie Caillat, "Bubbly")
  - *whole/home*; *close/home* × 2; *nine/life* × 2 (Emimem, "Lose Yourself")
  - *roof/moon* (Tom Petty, "Even the Losers")
- Seven features different: 1
  - *whole/broke* (Emimem, "Lose Yourself")

Place Features are to Blame. . .

- Low number of 1-feature differences: place features
- A multivalued [Place] feature smooths things out:



Multivalued Features Everywhere



Interim Summary

- Distinctive features do a good job of modeling imperfect rhyme frequency.

⇒ Featural differences match speakers' similarity intuitions. . .

- Except for place features: mismatches in place mean a large number of featural differences, but this is not reflected in the frequency of pairs mismatching in place.
- Fewer multivalued features perform better than many binary features.
- Next step: compare specific feature systems.

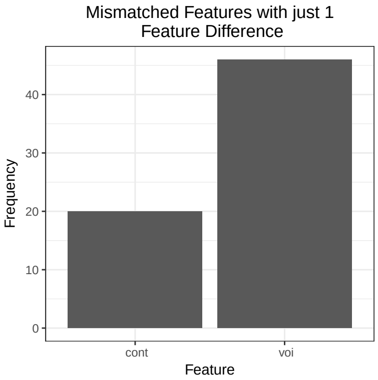
A Closer Look

Not All Features are Equal

- If consonants mismatch on exactly one feature:

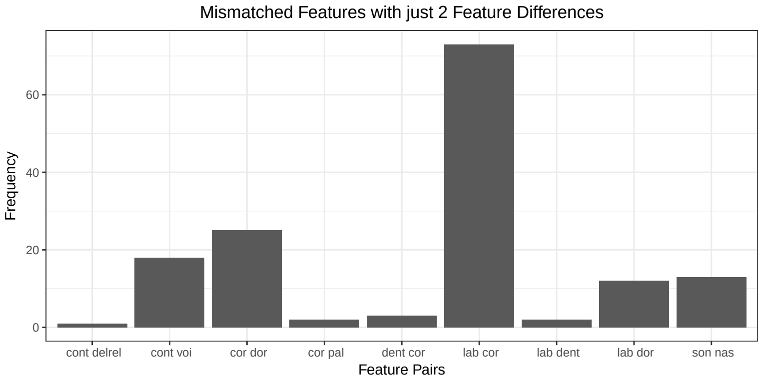
Not All Features are Equal

- If consonants mismatch on exactly one feature:



Not All Features are Equal

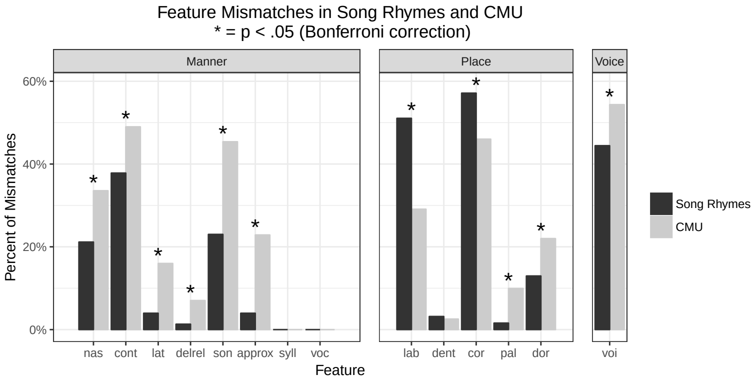
- If exactly two features mismatch:



Not All Features are Equal

- Some mismatch more than others.
- To ensure this isn't simply a reflection of consonantal frequency, we did the same analysis on the portion of the CMU dictionary that also occurs in CELEX (Baayen et al. 1995) to weed out low-frequency items:
  - Match each final-stress word to all other words with the same final vowel and same number of consonants
  - Compare coda consonants as before

Our Data vs. CMU/CELEX



- Over represented: [lab, cor]
- Under represented: nearly everything else
- Mismatches on [lab, cor] are more acceptable. Perhaps differences along these dimensions are "smaller" than differences along other dimensions.

What's up with [lab] & [cor]?

- [m]∼[n]: 31.1% (60/193) of all [lab] mismatches; 27.8% (60/216) of [cor] mismatches.
- This accounts entirely for the prevalence of [lab] and [cor] mismatches.
- We can't explain the high frequency of [m]∼[n] merely on the grounds that place cues for nasals are weak: why are [n]∼[ŋ] and [m]∼[ŋ] infrequent?
  - 9 tokens of [n]∼[ŋ]; 18.4% of [dor] mismatches, 4.2% of [cor] mismatches
  - 1 token of [m]∼[ŋ]; 2.0% of [dor] mismatches, .5% of [lab] mismatches
- It looks like a combination of nasal place weakness and a preference for [lab]/[cor].

Implications

What this might mean:

- Certain feature (mis)matches are more significant than others, as are certain combinations.
- E.g. labials and coronals are judged as more similar than, say, labials and dentals, stops and fricatives, etc.

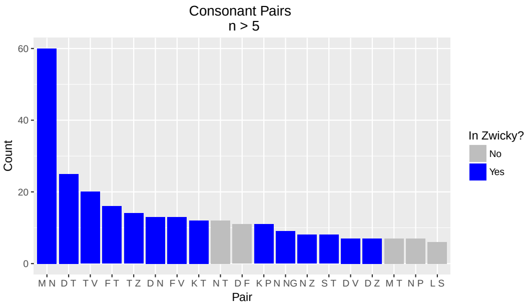
Implications

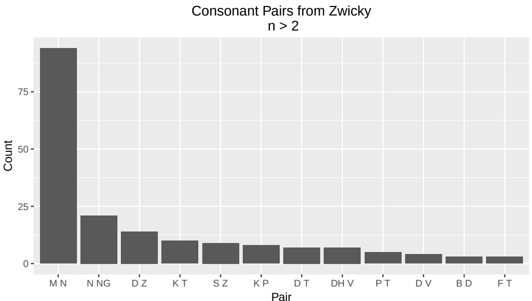
In phonological systems:

- If featural asymmetries matter to grammars, they should arise in the typology of ABC/OCP systems.
  - Cooccurrence of similar consonants is disfavored in $C_1C_2C_3$ Arabic roots. Frisch et al. (2004): all combinations of non-identical place features in C_1 and C_3 are over represented, but labial/dorsal combinations are less over represented than others.
  - Not so for C_1 and C_2 though
- But maybe grammars don't care about these asymmetries. Grammars are a step removed from phonetic detail in other ways.

Comparison with Zwicky (1976)

Most Common Consonant Pairs

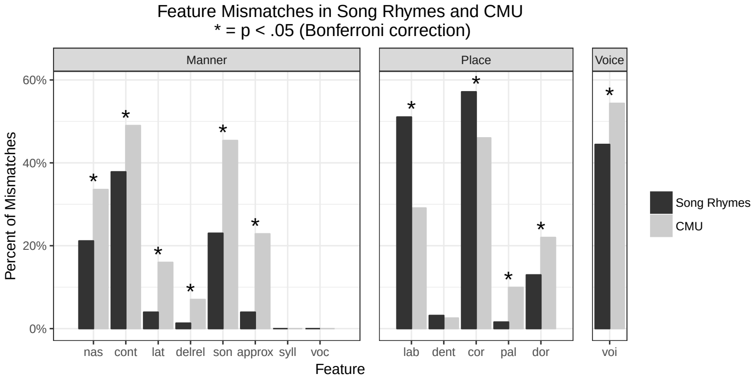




Most Common Consonant Pairs

- [m]∼[n] is the most common pair in both analyses, but:
  - It is 39.8% of all pairs in Zwicky
  - Only 15.9% in our data (60/378)
- [n]∼[ŋ] is second most common for Zwicky (8.9%)
  - 12th on our list (2.4%; 9/378)

Most Common Consonant Pairs



Zwicky's (1976) results (for feature mismatches $\geq$ 10):

- [dor] 148
- [lab] 138
- [cor] 70
- [cont] 49
- [voi] 19
- [pal] 10

Conclusion

Conclusion

- Generally, fewer featural differences between consonants makes them more likely to be paired in rhymes.
- Except for place features, counting features is a plausible model of speakers' similarity judgments.
- But the particular features involved matters, too: do some they represent smaller differences?

Next Steps

- Vowels
- Differences in number of consonants
- Compare specific feature systems
- Morphology (Zwicky 1976): e.g. does past-tense /d/ behave differently from other /d/?
- Genre & year differences

Baayen, Harald R., Richard Piepenbrock, & Leon Gulikers (1995) *CELEX2 LDC96L14*. Philadelphia: Linguistic Data Consortium.

Frisch, Stefan A., Janet B. Pierrehumbert, & Michael B. Broe (2004) Similarity Avoidance and the OCP. *NLLT* **22**: 179–228.

Hayes, Bruce (2009) *Introductory phonology*. Malden, MA: Wiley-Blackwell.

Leben, Will (1973) *Suprasegmental Phonology*. Ph.D. thesis, MIT.

McCarthy, John & Alan Prince (1995) Faithfulness and Reduplicative Identity. In *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, Jill Beckman, Laura Walsh Dickey, & Suzanne Urbanczyk, eds., 149–348, Amherst, MA: GLSA.

Rose, Sharon & Rachel Walker (2004) A Typology of Consonant Agreement as Correspondence. *Language* **80**(3): 475–531.

Zwicky, Arnold M. (1976) Well, this Rock and Roll has got to Stop. Junior's Head is Hard as a Rock. In *Papers from the 12th Regional Meeting, Chicago Linguistics Society*, Saliko S. Mufwene, Carol A. Walker, & Sanford B. Steever, eds., 676–697, Chicago: Chicago Linguistics Society.

Thanks to these students for their contributions to this work:

Maya Anderson-Badillo, Alayna Arnholt, Erin Bass, Jason Blaisdell, Wesley Block, Alexander Carruth, Henry Chapman, Alyssa Chen, Henry Chen, Shuxin Chen, Kathryn Danis, Audrey Dockendorf, Carla Escobar-Tomlienovich, Elizabeth Farmer, Avi Feldman, Myers Harbinson, Erin Harrill, Hope Hickman, Moshe Ikechukwu, Aunindya Jyoti, Joshua Kersey, Siyun Lee, Sophia Luo, Gabrielle Marushack, Joseph McCarty, Abigail McNaughton, Chelsea Middleton, Sumani Nunna, Gehao Pang, Meghana Patel, Cecilia Poston, Anna Price, Carrington Rodgers, Sofia Sanchez-Zarate, Sarah Schuhler, Yasharth Singh, John Stephens, Raphael Villamorovic, Janie Wang, Alexander Xiong, Amy Zhang, and Jacky Zhang.