# A new approach to calculating allophonic balance and functional load

**Naomi Gurevich[a], Zhenisbek Assylbekov[a], Maxat Tezekbayev[b], Heejin Kim[c], and Jonathan Washington[d]**

[a]Purdue University Fort Wayne, USA; [b]Nazarbayev University, Kazakhstan; [c]University of Illinois, USA; [d]Swarthmore College, USA

2025 LSA ANNUAL MEETING — JANUARY 9-12 — #LSA2025

## BACKGROUND

- **Purpose**: New method to measure
  - Allophonic balance
  - Functional load
- Relevant to speech perception and intelligibility

### Measuring Functional load

- Consequences of loss of lexical distinction:
  - Greater with higher functional load contrast
- With respect to allophones:
  - Positional distribution (**allophonic coverage**)
  - Frequency thereof (**allophonic balance**)

### Sound structure constrained by contrast

- Typologically, overly similar contrasts are avoided
- When alveolar stops are spirantized (Gurevich 2011)
  - [d] → [ɾ] (always)
  - but [d] → [ɾ] only when no /r/, else → [ð]
- Rare occurrence of oppositions between similar phonemes
  - Basque laminal /ʃ/: apical /s̺/
- Allophonic balance may **predict the functional consequences**
- Clinical implications for evaluation of intelligibility (Gurevich & Kim 2023)

## METHOD

- Defined a series of specific universal contexts for phones
  - By adjacency to C, V, or # (word boundary)
  - By word length & stress status (1σ, >1ó, >1ŏ)

### Automatic Processing

- Transcribe using CharsiuG2P (Zhu et al. 2022)
- Syllabify it (our own algorithm)
- Count occurrences of each phone in each context
- Normalize the counts
- Output in table
- source code: https://github.com/MaxatTezekbayev/allophonic_balance

### Application

- Can use a large corpus to represent a language
- Can compare individual texts to corpus results
  - For benchmarking

## Stop using *The Rainbow* passage!
### Ask me why.

## RESULTS

### Applying to English

- Used 188K-sentence IWSLT'14 corpus (Cettolo et al. 2014)
- E.g., in **1σ, #_V**:
  - /z/ is extremely uncommon
  - /ð/ is very common
    - Very frequent in function words, 2nd least frequent in content words (Gurevich & Kim 2022)
  - ➤ The z/ð contrast: **low functional load**
    - If neutralized would not greatly interfere with communication
  - /t/ and /d/ are both fairly common
  - ➤ The t/d contrast: **higher functional load**

### Benchmarking of a text:

- *The Rainbow Passage* (Fairbanks 1960)
  - 35.84% allophonic coverage (Gurevich & Kim 2022)
  - 86.34% allophonic balance
- This passage is not representative of English speech
  - Limits clinical and research usefulness of the passages
- Recommended alternative: FITI phrases (Gurevich & Kim 2023)

## CONCLUSIONS

### Problems with other methods

- Massive variability of stimuli across studies
- Hinders comparison of results (Bent et al. 2007)
- Allophonic balance ignored by tools to calculate functional load (Hall et al. 2021)

### Advantages of our method

- Semi-automated
- Cross-linguistic use
- Facilitates replication and comparison of results across studies

## References

Bent, T, Bradlow, A. R., & Smith, B. L. 2007. Phonemic errors in different word positions and their effects on intelligibility of non-native speech. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, 331–347. John Benjamins Publishing Co..

Cettolo, M., Niehues, J., Stüker, S., Bentivogli, L., and Federico, M. 2014. Report on the 11th IWSLT evaluation campaign. In *International Workshop on Spoken Language Translation* (IWSLT).

Fairbanks, G. 1960. *Voice and articulation drillbook*. Harper & Row.

Gurevich, N. 2011. Lenition. In M. Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 3). Wiley-Blackwell.

Gurevich, N., & Kim, H. 2022. Examination of consonantal phonetic coverage in standard reading passages. *Perspectives of the ASHA Special Interest Groups*, 7(5), 1573–1582. doi: 10.1044/2022_PERSP-22-00063

Gurevich, N., & Kim, H. 2023. Development of novel speech stimuli with phonetic coverage and phonemic balance. *Perspectives of the ASHA Special Interest Groups*, 8(2), 424-437. doi: 10.1044/2023_PERSP-22-00180

Hall, Kathleen Currie, Blake Allen, Edith Coates, Michael Fry, Serena Huang, Khia Johnson, Roger Lo, Scott Mackie, Stanley Nam, and Michael McAuliffe. (2021). *Phonological CorpusTools*, Version 1.5.0.. [Computer program]. http://phonologicalcorpustools.github.io/CorpusTools/

Zhu, J., C. Zhang, and D. Jurgens. 2022. "Byt5 model for massively multilingual grapheme-to-phoneme conversion." https://arxiv.org/abs/2204.03067

| /t/ | 0.0177952 |
| /d/ | 0.0056267 |
| /k/ | 0.0074241 |

| /ð/ | 0.0466904 |
| /s/ | 0.0110987 |
| /z/ | 5.869E-05 |

*Positional context counts of English part of IWSLT'14 corpus (normalized counts)*

| | #_V 1σ | #_V >1ó | #_V >1ŏ | #_C 1σ | #_C >1ó | #_C >1ŏ | V_# 1σ | V_# >1ó | V_# >1ŏ | C_# 1σ | C_# >1ó | C_# >1ŏ | V_C 1σ | V_C >1ó | V_C >1ŏ | C_V 1σ | C_V >1ó | C_V >1ŏ | V_V >1ó | V_V >1ŏ | C_C 1σ | C_C >1ó | C_C >1ŏ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| /p/ | 0.0023935 | 0.0051344 | 0.0014987 | 0.0009998 | 0.0027569 | 0.0010457 | 0.0025005 | 2.486E-05 | 0.0004248 | 0.0002909 | 6.25E-06 | 4.076E-06 | 0.000528 | 0.0010734 | 0.0010372 | 0.0006305 | 0.0021588 | 0.0016787 | 0.0009972 | 0.0004553 | 0.0002088 | 0.0002099 | 0.0005179 |
| /b/ | 0.0098783 | 0.0032224 | 0.0032801 | 0.0013021 | 0.0006123 | 6.453E-05 | 0.0005604 | 4.932E-05 | 3.329E-05 | 1.331E-05 | 7.744E-06 | 4.755E-06 | 0.0001192 | 0.0004668 | 0.0015123 | x | 0.0002087 | 0.001356 | 0.0028053 | 0.0031683 | 1.467E-05 | 3.736E-05 | 0.0001096 |
| /t/ | 0.0177952 | 0.0026534 | 0.0011248 | 0.0009739 | 0.0008521 | 0.0003732 | 0.0314937 | 0.0029277 | 0.0025981 | 0.0104114 | 0.0019316 | 0.0059218 | 0.0050865 | 0.0010949 | 0.0015677 | 0.0018052 | 0.0039484 | 0.0104117 | 0.0021666 | 0.0109199 | 0.0009589 | 0.0018147 | 0.0031163 |
| /d/ | 0.0056267 | 0.0038865 | 0.0031205 | 0.0004865 | 0.0001698 | 2.717E-05 | 0.0081531 | 0.000589 | 0.003884 | 0.0213797 | 0.0016851 | 0.002534 | 0.0005697 | 0.0004971 | 0.0004709 | x | 0.0004137 | 0.0035236 | 0.0017124 | 0.0049493 | 0.0007789 | 0.0004713 | 0.0007437 |
| /k/ | 0.0074241 | 0.0041187 | 0.0033123 | 0.0010706 | 0.0014282 | 0.0005193 | 0.0057837 | 0.0001859 | 0.0018067 | 0.0025868 | 1.44E-05 | 0.0001172 | 0.0021453 | 0.0054936 | 0.0052424 | 0.0005968 | 0.0005852 | 0.0017214 | 0.0029588 | 0.0044589 | 0.0007084 | 0.0006688 | 0.0004183 |
| /g/ | 0.0044632 | 0.0021928 | 6.005E-05 | 0.0012915 | 0.0006373 | 3.492E-05 | 0.0005542 | 3.125E-06 | 5.108E-05 | 3.125E-06 | 1.359E-06 | 3.6E-05 | 0.0001935 | 0.0006231 | 0.0012984 | 1.359E-07 | 0.0002137 | 0.0006544 | 0.0009884 | 0.0010149 | 1.902E-05 | 0.0001186 | 0.0003701 |
| /f/ | 0.0071548 | 0.002958 | 0.0005762 | 0.0026489 | 0.0006895 | 5.163E-05 | 0.0146906 | 0.0002232 | 9.864E-05 | 7.35E-05 | 0.0002258 | 0.0001216 | 0.000321 | 0.000722 | 0.0006174 | 1.617E-05 | 0.0003844 | 0.0013579 | 0.001043 | 0.0002969 | 4.076E-06 | 0.0001465 | 0.0002171 |
| /v/ | 0.0001793 | 0.0024962 | 0.0002471 | 8.722E-05 | 1.277E-06 | x | 0.0051734 | 0.0004338 | 0.0008119 | 9.538E-05 | 5.665E-05 | 9.51E-07 | 0.0004728 | 0.0004323 | 0.0011412 | 9.51E-07 | 0.0008747 | 0.0005237 | 0.0013579 | 0.0051375 | 6.521E-05 | 0.0003278 | 4.117E-05 |
| /θ/ | 0.0034432 | 0.0005159 | 2.255E-05 | 0.0008652 | 9.51E-05 | 1.549E-05 | 0.0030508 | 2.391E-05 | 6.168E-05 | 0.0005895 | 1.956E-05 | 3.288E-05 | 3.628E-05 | 5.489E-05 | 4.035E-06 | x | 5.679E-05 | 0.0008364 | 0.0002375 | 0.0009532 | 0.0001329 | 7.241E-05 | 4.538E-05 |
| /ð/ | 0.0466904 | 0.0005732 | 0.0001047 | - | - | - | 2.35E-05 | 1.359E-07 | 4.076E-07 | x | x | x | 5.57E-06 | 8.152E-06 | 6.793E-07 | x | 9.334E-05 | x | 0.0021208 | x | - | - | 3.804E-06 |
| /s/ | 0.0110987 | 0.0063134 | 0.0017941 | 0.004323 | 0.002902 | 0.0001769 | 0.015438 | 0.0007301 | 0.0026227 | 0.0087892 | 0.0007793 | 0.0040662 | 0.0041348 | 0.0024855 | 0.0050045 | 2.989E-06 | 0.0020411 | 0.0038499 | 0.0016727 | 0.0042752 | 0.0009929 | 0.0014041 | 0.0017896 |
| /z/ | 5.869E-05 | 0.0001069 | 1.209E-05 | 1.359E-07 | 2.717E-06 | 2.717E-07 | 0.0107491 | 0.0018343 | 0.0046059 | 0.0073472 | 0.000821 | 0.007419 | 0.0003142 | 0.0004341 | 0.0002759 | x | 0.0006365 | 0.0001087 | 0.0012018 | 0.0031584 | 5.434E-07 | 2.609E-05 | 6.657E-05 |
| /ʃ/ | 0.0021249 | 0.0003869 | 1.671E-05 | 2.568E-05 | 1.671E-05 | 4.076E-07 | 0.002765 | 4.212E-06 | 0.0002319 | 4.483E-06 | x | 1.359E-07 | 3.641E-05 | 5.04E-05 | 0.0001219 | 1.359E-07 | 2.473E-05 | 0.0018155 | 0.0001698 | 0.0051205 | - | 3.804E-06 | 3.668E-06 |
| /ʒ/ | - | - | - | - | - | - | 9.51E-07 | 1.128E-06 | 1.63E-06 | - | - | - | - | - | - | x | 1.766E-06 | 6.576E-05 | 1.25E-05 | 0.0005797 | - | - | - |
| /h/ | 0.013287 | 0.0027266 | 0.000313 | - | - | - | - | - | - | - | - | - | - | - | - | x | 0.0001047 | 0.0003751 | 0.0003166 | 0.0001466 | - | - | - |
| /tʃ/ | 0.0009455 | 0.000869 | 8.858E-05 | x | 4.076E-07 | - | 0.0024823 | 0.0001056 | 2.595E-05 | 0.0001827 | 0.000109 | 5.434E-06 | 6.494E-05 | 0.000185 | 6.657E-06 | x | 0.0001499 | 0.0024543 | 0.00061 | 0.001034 | 2.758E-05 | 1.196E-05 | 1.359E-06 |
| /dʒ/ | 0.0020296 | 0.000761 | 0.0003069 | 4.076E-07 | 1.359E-07 | 2.717E-07 | 0.0004 | 2.772E-05 | 0.0007274 | 0.0005187 | 4.375E-05 | 7.785E-05 | 1.196E-05 | 0.0004 | 9.13E-05 | x | 0.0001648 | 0.0011029 | 0.0002831 | 0.0024667 | 8.845E-05 | 5.38E-05 | 8.016E-06 |
| /m/ | 0.0093325 | 0.0053399 | 0.0010664 | 3.261E-06 | 0.0001515 | 7.201E-05 | 0.0079867 | 0.0004763 | 0.0017658 | 0.000317 | 5.801E-05 | 5.869E-05 | 0.0009207 | 0.0033771 | 0.0028466 | 0.0002888 | 0.0005876 | 0.0018761 | 0.0022326 | 0.0046159 | 0.0002224 | 8.057E-05 | 3.016E-05 |
| /n/ | 0.0077663 | 0.0023522 | 0.0002912 | 1.494E-06 | 2.717E-07 | 5.434E-07 | 0.023301 | 0.0015313 | 0.0103058 | 0.0003471 | 3.994E-05 | 0.000233 | 0.0231505 | 0.0121612 | 0.0163289 | 4.551E-05 | 0.0006547 | 0.0015611 | 0.0021651 | 0.0083067 | 0.0003467 | 4.864E-05 | 5.38E-05 |
| /ŋ/ | - | - | - | - | - | - | 0.0014801 | 0.0001456 | 0.0128486 | - | - | - | 0.0026004 | 0.0012228 | 0.000461 | - | - | - | 0.0002007 | 9.646E-06 | - | - | - |
| /w/ | 0.0274076 | 0.0026568 | 0.0003128 | - | - | - | - | - | - | - | - | - | 0.0029111 | 0.001363 | 0.001513 | 0.0003667 | 0.0006665 | - | - | - | - | - | - |
| /j/ | 0.0107154 | 0.0010225 | 0.0004269 | - | - | - | - | - | - | - | - | - | 0.0005251 | 0.0018415 | 0.0029138 | 0.0001575 | 8.369E-05 | - | - | - | - | - | - |
| /l/ | 0.0075149 | 0.00351 | 0.0001531 | - | - | - | 0.0073332 | 0.0003736 | 0.0099454 | 7.513E-05 | 4.076E-07 | 8.152E-07 | 0.0032411 | 0.0044674 | 0.002477 | 0.0022587 | 0.0028322 | 0.0056669 | 0.0028852 | 0.0095182 | 0.0009709 | 4.375E-05 | 5.978E-06 |
| /r/ | 0.0028363 | 0.0033095 | 0.0027551 | - | - | - | 0.0184139 | 0.0008006 | 0.0134525 | - | - | - | 0.0090718 | 0.0082023 | 0.0090589 | 0.0072915 | 0.0073936 | 0.0075908 | 0.0004842 | 0.0110974 | - | - | - |