Quantifying Phonetic Informativity: An Information Theoretic Approach

James Whang

Seoul National University (Korea)
jamesw@snu.ac.kr

For successful speech perception, a listener must assign higher weights to phonetic cues that are crucial for categorical contrasts than to non-crucial ones. The problem, however, is that phonetic cues are often redundant in natural speech [1], making weighting a non-trivial task. It is expected then that there are individual differences in how specific cues are weighted which in turn can lead to language change over time. Indeed such diachronic 'mis'-weighting is considered a major driver of language change [2]. In information theoretic [3] terms, redundant cues have low informativity. Research in segmental [4] and lexical [5] domains show that it is precisely such low information units that are targeted for reduction. The current study presents the results of a working model that extends this information theoretic approach to the subsegmental level.

The phonetic informativity quantification model presented here uses surprisal to quantify redundancy and entropy to quantify overall information of a given feature x.

Surprisal: $$-\log_2 \Pr(x|Context) \quad (1)$$
Entropy (H): $$\sum \Pr(x|Context) * -\log_2 \Pr(x|Context) \quad (2)$$

Given the formulation above, zero surprisal for x would mean that it is completely redundant in the context of one or more other cues. Zero entropy would mean that x is completely uninformative and has no function in the language. Both surprisal and entropy have no theoretical upper bound, but the higher the value, the higher the functional load in the language.

As a test case, the current study focuses on the informativity of vowel cues in Japanese. Vowel features are used as stand-ins for acoustic cues for simplicity's sake in the current study. All calculations are based on CSJ-RDB, a 500K-word subset of the Corpus of Spontaneous Japanese [6]. The corpus contains a total of 679,123 vowels. The model first converts each vowel to a user-defined, n-sized set of phonetic features, with each feature's frequency being equal to the sum of the vowel frequencies that contained the feature. Second, surprisal (redundancy) is calculated for each feature with all possible subsets of the remaining features as context, resulting in 2^{n-1} contexts. For example, given a vowel with feature set $A = \{a, b, c, d\}$, surprisal for feature a is calculated with all possible subsets of $A \setminus a$ as context (including {} and $\{b, c, d\}$), then again for b with all possible subsets of $A \setminus b$, etc. Lastly, entropy (informativity) is calculated based on the surprisal values.

For the current analysis, the set consisted of the following eight features with their respective parameters shown in []:

V → {**height** [*high, mid, low*], **backness** [*front, central, back*], **roundedness** [*rounded, unrounded*], **length** [*short, long*], **peripherality** [*centralized, peripheral*], Δ**height** [*level, rising, falling*], Δ**backness** [*stable, fronting, backing*], Δ**roundedess** [*constant, rounding, unrounding*]}

Although the CSJ-RDB annotations only contained short and long monophthongs (/i, e, a, o, u, ii, ee, aa, oo, uu/), the three delta features were included under the assumption that the set of possible features must be the same across all languages. All Japanese vowels, therefore, were {…, *level, stable, constant*} but this would not be case for languages that have vowels with significant movement (e.g., English /aʊ/ → {…, *rising, backing, rounding*}).

Three key results are reported here in the interests of space. First as expected all delta features were completely redundant in Japanese, with zero surprisal in all contexts and consequently zero entropy. The lack of informativity for these features suggests that Japanese listeners are insensitive to vowel movement. Second, the feature [*high*] had the highest entropy (97.80), suggesting that

Japanese listeners have heightened sensitivity to high vowels. Lastly, [*long*] had the second highest entropy (88.58). In conjunction with [*peripheral*] having zero entropy, the results predict that Japanese listeners should be more sensitive to length manipulations than peripherality manipulations. All three predictions are supported by previous experimental studies [7,8,9].

The current study presents a simple working model for quantifying the informativity of phonetic cues/features in a given language. The use of Information Theory places the model in a long line of research that sought to understand gradient predictability effects in phonetics and phonology [4,5,8,10]. It also shows the theory's incredible flexibility to handle linguistic representations of varying sizes and granularity. Future work includes further developing the model to process not just simple sets of features but also sequences of cue/feature vectors to quantify segment-internal timing relations [11], as well as extending the model to fine-tune cross-linguistic perception frameworks [12,13], which typically rely on cue-weighting differences to explain perceptual errors.

References

[1] Clements, G. N. (2009). The role of features in phonological inventories. *Contemporary views on architecture and representations in phonological theory*, 19-68.

[2] Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

[3] Shannon, C. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

[4] Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2). 243–278.

[5] Hall, K. C., Hume, E., Jaeger, F. T., & Wedel, A. (2016). The message shapes phonology. Ms. UBC, University of Canterbury, University of Rochester and University of Arizona.

[6] Maekawa, K., and Kikuchi, H. (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In *Voicing in Japanese*, ed. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Mouton de Gruyter.

[7] Strange. W, Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Soceity of America, 109,* 1691-1704.

[8] Whang, J. (2019). Effects of phonotactic predictability on sensitivity to phonetic detail. *Laboratory Phonology*, *10*(1):8.

[9] Strange, W., Hisagi, M., Akahane-Yamada, R., Kubo, R. (2011). Cross-language perceptual similarity predicts categorial discrimination of American vowels by naïve Japanese listeners. *The Journal of the Acoustical Society of America* 130: EL226–31.

[10] Cherry, E. C., Halle, M., & Jakobson, R. (1953). Toward the logical description of languages in their phonemic aspect. *Language*, 34-46.

[11] Inkelas, S., & Shih, S. S. (2017, May). Looking into segments. In *Proceedings of the Annual Meetings on Phonology* (Vol. 4).

[12] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233–77.

[13] Best, C. T. (1995). A direct realist view of cross-language speech perception. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171–204.