Generative Phonetic Modeling of Malay Nasal Harmony

Background: Malay (Austronesian) exhibits nasal harmony whereby nasality spreads iteratively and rightward from nasal consonants to vowels and glides, but is blocked by supralaryngeal consonants (e.g., /mewah/ → [mẽw̃ãh̃], but /makan/ → [mãkan] *[mãkãn]). A key puzzle arises in reduplicated forms where Onn (1976:69-70) reports overapplication of nasal harmony (e.g., [**w̃ãŋĩ**-w̃ãŋĩ]): the first syllable of the reduplicant (the underlined constituent) acquires nasality even though there is no trigger preceding it. This pattern has significant theoretical importance because only parallelist (McCarthy & Prince 1995) but not serial/derivational theories of reduplication (Inkelas & Zoll 2005, Kiparsky 2010, McCarthy *et al*. 2012) can account for this pattern.

Methodology: The present study reports acoustic data collected in person with 30 native speakers of Malay. Subjects read 18 Malay target words in isolation (e.g., *wa*$_{I}$*ngi*) once and in reduplicated forms (e.g., *wa*$_{R}$*ngi-wa*$_{B}$*ngi*) three times in randomized order. All words were embedded in a carrier sentence that contained no nasal segments. The nasality of vowels in the reduplicant (a_R) and the base (a_R) was assessed using the A1−P0 acoustic measure (Chen 1997; Styler 2017), normalized per speaker and per word using the vowel in isolation form (a_I) as the oral baseline.

Results: Our results shown in Figure 1 confirm the overapplication pattern reported in Onn (1976), but it exists as one of four variable outputs, with overapplication being the most frequent output (Quadrant I: 39.5%), followed by underapplication ([waŋĩ-waŋĩ], Quadrant III: 27.5%), normal application ([waŋĩ-w̃ãŋĩ], Quadrant IV: 24.4%) and pathological application ([w̃ãŋĩ-waŋĩ], Quadrant II: 8.6%). The results also display several interesting gradient patterns: (i) most of the data points clustered around the origin, i.e., the oral baseline (the heat map shown in Figure 2) and (ii) the vowels of reduplicants and bases tend to maintain the same acoustic level of orality/nasality.

Modeling simulation: In order to capture the variable and gradient patterns seen in the observed data, we developed a constraint-based model couched in Maximum Entropy Harmonic Grammar (MaxEnt), building on McCarthy & Prince (1995:42)'s analysis of Malay nasal harmony in classical Optimality Theory. There are two crucial differences between their analysis and ours. First, instead of using strict ranking of constraints, we employed constraint weighting to derive the variability seen in the outputs of nasal harmony. Second, instead of assessing constraint violations categorically, violations were computed numerically along a continuous scale, following Flemming (2001) and Lefkowitz (2017). The simulation was done using the *maxent.ot* package in R (Mayer *et al.* 2024) and the model's predictions yielded a good overall fit to the observed data (Figure 3). The constraints employed in the MaxEnt model and their relative weights are shown in Figure 4.

Discussion: The existence of overapplication of nasal harmony in Malay bolsters the argument for parallelist theories of reduplication. Moreover, the MaxEnt learning simulation demonstrates that both gradient and variable patterns can be modeled within a unified, constraint-based framework.

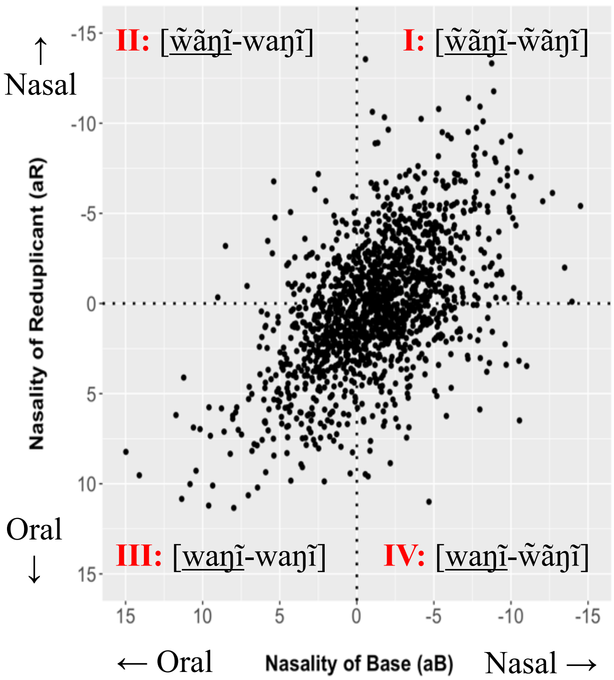

Figure 1. Results from the acoustic study

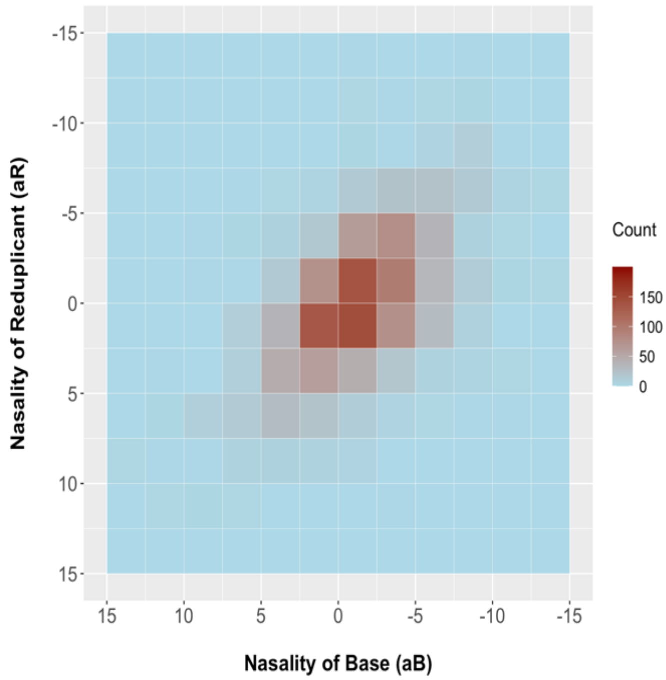

Figure 2. Heat map of the acoustic results

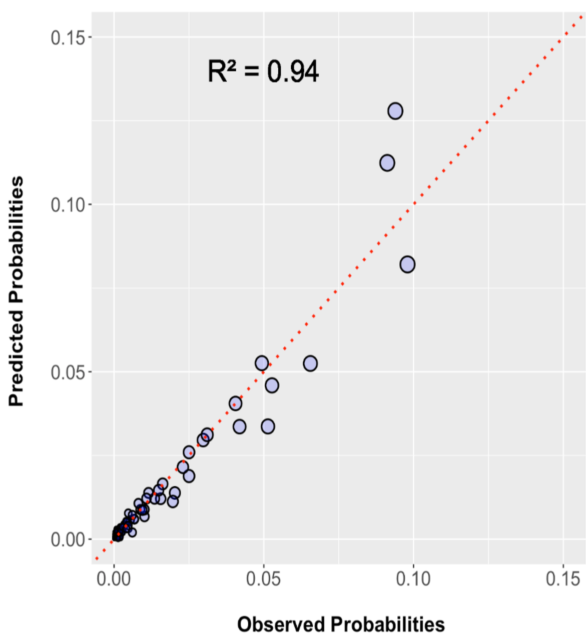

Figure 3. Observed vs. predicted probabilities

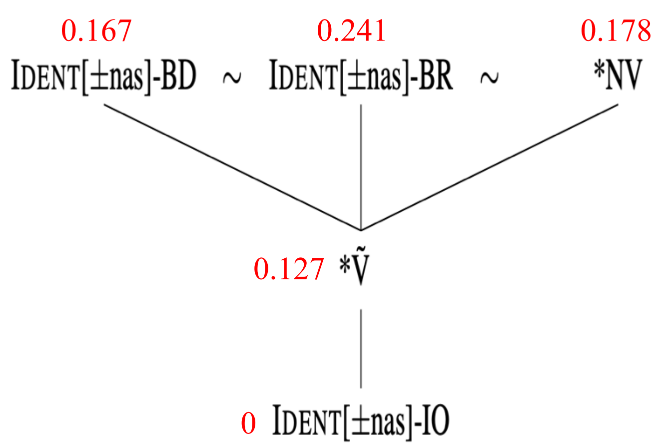

Figure 4. Constraints and their relative weights