Language Learning and Development, 7: 243–249, 2011

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2011.578547

Psychology Press
Taylor & Francis Group

RESPONSE

Response to MacNeilage and Davis and to Oller

D. H. Whalen

Haskins Laboratories, New Haven, Connecticut and Department of Speech–Language–Hearing Sciences, City University of New York

Sara Giulivi

Haskins Laboratories, New Haven, Connecticut and Department of Linguistics and Oriental Studies, University of Bologna

Louis M. Goldstein

Haskins Laboratories, New Haven, Connecticut and Department of Linguistics, University of Southern California

Hosung Nam

Haskins Laboratories, New Haven, Connecticut

Andrea G. Levitt

Haskins Laboratories, New Haven, Connecticut and Department of French, Wellesley College

The article by MacNeilage and Davis in this issue, entitled "In Defense of the 'Frames, then Content' (FC) Perspective on Speech Acquisition: A Response to Two Critiques" appears to assume that the only alternative to segment-level control is oscillation specifically of the jaw; however, other articulators could be oscillated by infants as well. This allows the preferred CV combinations to emerge without positing a level of segmental control in babbling. Their response does not address our modeling work, which, rather similarly to Davis's own modeling (Serkhane, Schwartz, Boë, Davis, & Matyear, 2007), shows little support for the Frame-then-Content (F/C) account. Our results show

Correspondence should be addressed to D. H. Whalen, Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511. E-mail: whalen@haskins.yale.edu

substantial support for the Articulatory Phonology (AP) one. A closer look at feeding in infants shows substantial control of the tongue and lips, casting further doubt on the foundation of the F/C account.

In this response, we would like to address the issues raised in two commentaries, which were prompted by our article, "An Articulatory Phonology Account of Preferred Consonant-Vowel Combinations" (2011, this issue). We will first respond to "In Defense of the "Frames, then Content (FC) Perspective on Speech Acquisition: A Response to Two Critiques," by Peter F. MacNeilage and Barbara L. Davis; we will then turn to "Vocal Category Development in Human Infancy: A Commentary on Giulivi et al.'s Critique of the Frames, then Content Model," by D. Kimbrough Oller.

MACNEILAGE AND DAVIS

MacNeilage and Davis are committed to the view that there is a lack of segmental-level control throughout the babbling period. One quote, among many possible, comes from MacNeilage (2008, p. 121): " . . . there is little evidence for segmental independence in the first year of speech-like production—the babbling and first word stages. There may not be any segmental independence in babbling; that is, infants may not have systematic control of single segments of the kind that enables them to insert one into various contexts." Segmental independence means the ability to combine an arbitrary action of one articulator with one arbitrary action of another articulator at a segmental time scale. They take as evidence for this both the CV ratios and also the inter-syllabic constraints on successive syllables. As they note in their response, "This result implies a virtually absent ability (or even a totally absent ability, given transcription uncertainties) to use the tongue to differentiate output across syllables" (MacNeilage & Davis, 2011, p. 238) and then, "If there is little ability to use articulators other than the mandible to vary patterns across syllables, how much versatility should we expect of these organs in varying these patterns within syllables?" (p. 238). Their alternative to segmental control is clearly "frame dominance" oscillation of the jaw. However, just because infants cannot produce arbitrary combinations of articulator actions at a segment-level time scale does not mean they are only controlling the jaw. It is perfectly possible (and indeed in our view more plausible) that infants are actively controlling (in an oscillating fashion) articulators other than the jaw, separately or in combination. We would expect that, in the case of multiple oscillating articulators, the oscillators would be in-phase, and indeed such in-phase coupling is basis for the synergy simulations in the AP model that make superior predictions about the CV ratios (both on- and off-diagonal). So the issue is not whether infants have a freely combinatorial segment system in the first year of life (neither theory proposes that) but rather whether oscillation is restricted to the mandible or whether other articulators can also be rhythmically engaged, separately and in combination. The AP simulations reported in the target article suggest that they can be.

MacNeilage and Davis further assert that "no evidence is provided for the assumption implicit in this work that adult speech and the initial babbling of infants have identical gestures and intergestural relationships" (p. 235). That is because we explicitly do not make that assumption. We assume that infants begin with gestures that are globally related to adult gestures, but which

are explicitly not timed the way that adult productions are. Mastery of the timing relationships is a large part of phonological development. We assume that various articulators are oscillating, not just the jaw, and that gestures that are more compatible with each other will be more likely to co-occur. MacNeilage and Davis claim that we are discussing high vowels only because we "even use these high vowels in capitalized notations for the presentation of the syllable-level fronting and backing tendencies in their infant data ("DI" and "GU")." The use of high vowels in an abbreviation (Giulivi et al., 2011) does not imply that we were dealing only with high vowels. As stated, the "I" stood for front vowels as low as [æ] and "U", for back vowels as low as [ɑ].

The basic premise of the F/C account is that only the mandible is controlled in both ingestion (at least in early evolution) and babbling, but human infant suckling is not solely based on jaw movement (and presumably never was); it includes active use of the lips and tongue as well. If the infant is not able to make a seal around the nipple with the lips, then there will not be enough suction for feeding to be possible (Ardran, Kemp, & Lind, 1958, p. 161). When milk is drawn into the mouth, the jaw typically moves downward, as assumed by the F/C model, but the tongue acts independently as well. As Riordan (2005, p. 86) said, "A rhythm is created by this sequence of vertical jaw movements and the depression and elevation of the posterior tongue." Thus the major feature of the "pure frame" stage—a lack of control of the articulators other than the tongue—is not true of feeding. The F/C account, then, requires us to believe that infants relinquish control over the tongue and lips in order to babble. This is inconsistent with both ontogeny and phylogeny.

The frequencies implied by the ratios of preferred CVs indicate that only about half of the syllables are in the predicted combinations. Indeed, Oller supports and extends this criticism in point three of his response. The acoustic measurements of Matyear et al. (1998) are taken to show that acoustic measurements would give higher ratios for the preferred combinations than do transcriptions, but the differences in the measured formants are small. Taking values from their Table 1, the average difference in F2 (the main acoustic measurement associated with the front/back dimension) is 449 Hz, while the average difference between their preferred and nonpreferred syllables is 42 Hz. Although positive and thus in the direction those authors predict, it is not clear whether this difference is large enough to move the syllable into the range of those that could be produced with jaw-only control, or indeed how this could be evaluated. As shown in our paper, non-diagonals can also be generated with jaw-only control (close to 50% of the time). The issue is how to account for the relative distribution across all the cells and the AP account appears to do better at this than the F/C account. Ultimately, the best way to resolve these issues will be to model the formant patterns of babbling directly, using the kinds of models we have tested (and possibly others). But adequate modeling and data are difficult to achieve, and for now, at least, it appears that the AP fit to the data is superior.

MacNeilage and Davis assert that "articulatorily synergetic factors they refer to must, if they actually exist, be universal" (p. 237) Although this is technically true, it is clear that patterns that affect less than half of the data will be less robust and thus more likely to appear inconsistently in any particular data set. Finally, while AP does assume that an adult language deploys a fixed number of gestural primitives, the discussion of how many there are, or their specific properties, is completely irrelevant to the issues being addressed here. The infants are not assumed to have adult gestures, or their organization, at this stage. The relevant issue is whether there is any independent control of the multiple articulatory degrees of freedom that will eventually develop into constriction "devices" (effectors) controlled by different gestures.

With respect to the relationship between mastication and jaw oscillation, MacNeilage and Davis single out a study by Hiiemae and Palmer (2003, p. 431) that "could argue in favor of an hypothesis which proposes that the movements of speech are a subset of those used in feeding" (p. 240), presumably leaving us to conclude that they must have done so. Even granting this subset relationship, this would not argue for jaw-only control, because as argued above, multiple articulators are coordinated in feeding. The relationship between speech and mastication is confusing in the F/C account. The relationship is phylogenetic only, but still "ontogeny recapitulates phylogeny in the sense of beginning with a similar set of motor constraints to those of early phylogeny, and then progressively acquiring the sounds and sound patterns" (MacNeilage & Davis, 2000, p. 285). It seems that mastication is, and is not, the source of babbled utterances in the F/C account. To the extent that the AP account makes any evolutionary assumptions, it is that speech, like other biological systems, is differentiated early in phylogeny and ontogeny: As soon as there is a functional difference, we can expect there to be a neurological difference that mitigates whatever similarities might have once existed between chewing and speech production.

While MacNeilage and Davis claim that there should be no developmental trend in the ratios during babbling and early words, they indicate that there should, after all, be a decline in the magnitude of the ratios at some point in time: "This clearly indicates the strength of the preferences for the three patterns is lower in languages than in infants" (p. 239). There is no other measure of "strength" besides the magnitude of the ratios. It thus appears that MacNeilage and Davis do, after all, acknowledge that there should be a decline in the size of the ratios as control of the articulators is gained. The fact that the ratios did not differ for first words and babbling is simply a fact that needs to be dealt with, not a prediction of the F/C account. The F/C account still requires a decline as articulation is mastered; otherwise, there is no explanation for the smaller effects in adult dictionary counts. The lack of such a decline during babbling is just what the target paper tested: We found there was no decrease in strength, and our results are consistent with unpublished results that MacNeilage and Davis mention in their response. Thus, our conclusion in the target paper is confirmed not only by our own results but also by those of MacNeilage and Davis. Apparently, MacNeilage and Davis hypothesize that the decline comes later, perhaps when children develop true segment-level combinatorial behavior. This would be consistent with their conflation of the issues discussed earlier: Jaw oscillation is the only alternative to segmental-level control. Surely we would expect some increased articulatory control in he first year of life, given other evidence for the overall increase in the maturation of infant utterances (Boysson-Bardies, 1999; Oller, 2000) during this period, as well as the appearance of "late" segments (Gildersleeve-Neumann, Davis, & MacNeilage, 2000). However, the appearance of these complex articulations does lead to a clear expectation that there should be a developmental trend in the CV ratios. The absence of such a trend, in our data as well as in that of MacNeilage and Davis, is not compatible with the F/C account.

MacNeilage and Davis assert that we adopt "two contradictory stances" (p. 239) on the ratios in adult language, but instead we point out that there are two contradictory results in adult language that need to be addressed and that MacNeilage and Davis do not acknowledge: Type counts and token counts differ. Although our target paper refers to the dictionary data for only the three languages we studied, we reach the same conclusions about the dictionary data that MacNeilage and Davis do for a larger set. The Mandarin dictionary results are not fully supportive of this conclusion, with only one out of three diagonals above one, but that same one-out-of-three pattern occurs for the German dictionary data (MacNeilage, Davis, Kinney, & Matyear, 2000). Structural

effects in Mandarin (affrication of alveolars before high front vowels) are likely to have affected these counts. We agree with MacNeilage and Davis that dictionaries show a (weak) preference for the same syllables. It is unclear why a mechanical linkage to the jaw that is overcome in development should continue to influence adult language, even if only in the creation of new words. For AP, the same synergies that predict the infant pattern are present in the adult language, and so could exert an influence on the selection of new lexical items. Moreover, MacNeilage and Davis take the lower ratios for the preferred syllables to be evidence for ultimate escape from frame dominance, but it is not clear how the statistical reliability of this decline could be established. Token counts, that is, assessments of CV co-occurrence in spoken corpora, indicate that French resembles the dictionary counts, Mandarin has two out of three diagonals above 1, and English has one out of three diagonals above 1 (Whalen et al., submitted). Indeed, the one that English has is the one that Mandarin lacks. The babbling results match the dictionary counts for all three languages, but the token (spoken corpus) counts only for French. Thus, there are two aspects of adult language that are in conflict, requiring different explanations. There is no contradiction in our account.

The intersyllabic effects that are discussed in MacNeilage and Davis (p. 237–238) do show that infants are not selecting arbitrary combinations of articulations in adjacent syllables, but rather they are engaged in a kind of oscillation. However, as noted at the outset here, evidence for oscillation is not by itself evidence that only the jaw is oscillating. Other articulators can, and we think are, being rhythmically engaged. It is hardly surprising that an oscillator operating for one syllable would continue to operate for successive syllables (that is what it means to oscillate). The intersyllabic effects do not differentiate the accounts.

The paragraph discussing whether off-diagonal syllables represent active control (p. 237) is based on an earlier version of the target paper; MacNeilage and Davis did not have access to the change in the text. The modeling results show that even jaw motion alone will generate about 50% off-diagonal syllables, so the criticism of our former statement is correct. However, the reformulation makes it clear that the original impetus for F/C, the generation of purely diagonal CVs by jaw motion alone, is incorrect.

Calling early speech articulator movements "gestures" does not imply that they are fully developed, adult forms of speech. It does imply that there is a biological predisposition to use articulators in communication. The ubiquity of speech, arising without instruction in all but the most neurologically atypical, is consistent with an early start to the language learning process. Babbling appears to exploit that capability, exhibiting its characteristics from the earliest stages onward.

OLLER

Oller's commentary (2011, this issue) is generally supportive of our claims, with one primary caveat. He points out that the increase in the number of segments produced does not necessarily indicate an increase in control because the total number of syllables was larger in the later samples as well. However, the overall increase in control exhibited by other results in the literature (Boysson-Bardies, 1999; Gildersleeve-Neumann et al., 2000; Oller, 2000) is consistent with our interpretation. Further, there is an implication that the infants may have controlled those segments even earlier. Thus such an argument may weaken the case for developmental change, but

it makes an even stronger case that there is control of articulation from the beginning of babbling. In either case, the FC account is inconsistent with our results.

The issue of whether the relatively high frequency of front vowels contradicts the FC account does not seem convincing. MacNeilage and Davis's original rationale for using the observed to expected ratios still seems valid: The point of using such ratios is to overcome just the sort of intrinsic frequency differences that Oller points out. It does not matter for the synergy phenomenon if the categories have different inherent frequencies. Only the combinations matter.

Oller claims that transcriptions are a poor indication of the infant's vocal range, giving the example of a child that uses 50 distinct syllables to represent just 10 words. However, only the produced phonetic utterances are at issue with the synergy phenomenon, reducing the main concern back to the reliability of transcriptions per se. Indeed, they are not terribly reliable, as Oller points out, but it does not appear that particular acoustic measurements are sufficiently well motivated to replace those transcriptions. If nothing else, the appearance of the CV preferences across different laboratories gives us some sense of the replicability of the results. Our own perceptual studies (Giulivi, 2007) indicate that there is no observable bias toward hearing the preferred combinations, even in babbled utterances.

The importance of other developments in the prebabbling stage highlights, as Oller indicates, that there are many more aspects of speech emerging in this time frame than are considered in the FC account. However, the AP account does not make specific claims about these issues either. We can only acknowledge that the development of phonatory strategies is consistent with a broad and early push toward speech.

SUMMARY

The preferences for certain CV patterns, as discovered by MacNeilage and Davis, are consistently found and are of inherent interest. We have provided an alternative account based on the biomechanical preferences within an Articulatory Phonology framework that provides a better prediction of the babbling and first word results; it is not surprising that relatively weak biomechanical preferences would influence babbling more than adult language, given that the system is just getting organized. Similarly, it is not surprising that preferred syllables would be chosen more frequently in first words. Proposing a single mechanism allows us to understand how the preferences could continue into the lexicons of adult speech.

REFERENCES

Ardran, G. M., Kemp, F. H., & Lind, J. (1958). A cineradiographic study of breast feeding. *British Journal of Radiology*, *31*, 156–162.

Boysson-Bardies, B. de. (1999). *How language comes to children: From birth to two years* (M. B. DeBevoise, Trans.). Cambridge, MA: MIT Press.

Gildersleeve-Neumann, C. E., Davis, B. L., & MacNeilage, P. F. (2000). Contingencies governing the production of fricatives, affricates, and liquids in babbling. *Applied Psycholinguistics*, *21*, 341–363.

Giulivi, S. (2007). Vowels and consonants favored co-occurrences in language development. (Unpublished doctoral dissertation). University of Florence, Italy.

MacNeilage, P. F. (2008). *The origin of speech*. Oxford, England: Oxford University Press.

MacNeilage, P. F., & Davis, B. L. (2000). Deriving speech from nonspeech: A view from ontogeny. *Phonetica*, *57*, 284–296.

MacNeilage, P. F., & Davis, B. L. (2011, this issue). In Defense of the "Frames, then content" (FC) perspective on speech acquisition: A response to two critiques. *Language Learning and Development*, *7*, 234–242.

MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (1999). Origin of serial output complexity in speech. *Psychological Science*, *10*, 459–460.

MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, *71*, 153–163.

Matyear, C. L., MacNeilage, P. F., & Davis, B. L. (1998). Nasalization of vowels in nasal environments in babbling: Evidence for frame dominance. *Phonetica*, *55*, 1–17.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller, D. K. (2011, this issue). Vocal category development in human infancy: A commentary on Giulivi et al.'s critique of the Frames, then Content model. *Language Learning and Development*, *7*, 226–233.

Riordan, J. (Ed.). (2005). *Breastfeeding and human lactation* (3rd ed.). Sudbury, MA: Jones and Bartlett.

Serkhane, J. E., Schwartz, J.-L., Boë L.-J., Davis, B. L., & Matyear, C. L. (2007). Infants' vocalizations analyzed with an articulatory model: A preliminary report. *Journal of Phonetics*, *35*, 321–340.

Whalen, D. H., Giulivi, S., Nam, H., Levitt, A. G., Hallé, P. A., & Goldstein, L. M. (submitted). Biomechanically preferred consonant-vowel combinations occur in adult lexicons but not in spoken language.