

An Articulatory Phonology Account of Preferred Consonant-Vowel Combinations

Sara Giulivi

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

D. H. Whalen

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

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

Certain consonant/vowel combinations (labial/central, coronal/front, velar/back) are more frequent in babbling as well as, to a lesser extent, in adult language than chance would dictate. The “Frame then Content” (F/C) hypothesis (Davis & MacNeilage, 1994) attributes this pattern to biomechanical vocal-tract biases that change as infants mature. Articulatory Phonology (AP; Browman & Goldstein, 1989) attributes preferences to demands placed on shared articulators. F/C implies that preferences will diminish as articulatory control increases, while AP does not. Here, babbling from children at 6, 9, and 12 months in English, French, and Mandarin environments was examined. There was no developmental trend in CV preferences, although older ages exhibited greater articulatory control. A perception test showed no evidence of bias toward hearing the preferred combinations. Modeling using articulatory synthesis found limited support for F/C but more for AP, including data not originally encompassed in F/C. AP thus provides an alternative biomechanical explanation.

INTRODUCTION

Babbling has been a source of evidence about language acquisition, but it is also a source of many puzzles. The issues concerning the nature of babbling and the way it relates to speech are still intriguing and open to discussion. Is babbling guided by innate linguistic factors or entirely by biomechanical factors? It is clear that biomechanical factors must be involved, given that the vocal tract is a biomechanical system. So, is it also possible to attribute language-like features to babbling? Further, are the biomechanical influences on babbling related to the production of speech *per se*, or do they derive from movement patterns shared with other functions, such as feeding?

Jakobson (1968) proposed that babbling represents a period in which the entire repertoire of spoken language is randomly explored in a sort of “purposeless egocentric soliloquy of the child” (Jakobson, 1968, p. 22). In his view, babbling has nothing in common with early meaningful speech that reflects, instead, universal markedness relations in the progressive differentiation of a series of phonological oppositions. In conflict with Jakobson, Locke (1983) proposed the first explicitly biological approach to phonological acquisition, according to which the articulatory gestures produced by the child exhibit innate universal characteristics, regardless of whether they belong to babbling or first words. The babbling repertoire is seen as a universal set of sounds from which first words can emerge. Thus, while Jakobson assumes that universality reflects the workings of distinctive features, Locke (1993) posits a greater influence of biomechanical limitations. Others find that the evidence is too limited to make a firm decision (e.g., Oller, 2000). Even knowing whether the “segments” of babbling have any theoretical standing is debatable. Transcribing babbling with segmental equivalents does not imply that infants have segmental targets. Adult listeners, whether parents or researchers, react to babbling as being similar to language and thus transcribable. However, what the child is controlling in the babbled utterances is still an open question.

Other evidence has shown target language influences on babbling in more narrow analyses (e.g., Oller, Wieman, Doyle, & Ross, 1976). In one study, adult speakers of French were able to identify whether 8-month-old babblers were learning French or not (Boysson-Bardies, Sagart, & Durand, 1984). The vowel space, as defined acoustically, used in babbling is influenced by the space of the ambient language (Boysson-Bardies, Hallé, Sagart, & Durand, 1989). One study of four languages revealed differences in the segmental inventories (Boysson-Bardies & Vihman, 1991), despite earlier reports of a lack of effect. Intonational differences can be found as early as six months of age (Whalen, Levitt, & Wang, 1991). In all these cases, adaptation to the target language could be couched in terms of featural adaptation or sensorimotor resonance.

Common tendencies for the relative frequency of certain syllable structures have been found in babbling and dictionary entries. Most babbling includes CV syllables, while consonant clusters and final consonants are rare (e.g., Oller, 2000). Here as well, the patterns could be explained in either phonological or biomechanical/dynamical frameworks. The unmarked nature of CV syllables across languages is quite clear (despite the possibly exceptional case of the Australian language known as Arrernte; Breen & Pensalfini, 1999). However, being consistent with a posited linguistic universal does not ensure that the phonological account is correct; other constraints may also play a role. Nam, Goldstein, and Saltzman (2009) show that the CV preference emerges in a simple Hebbian learning model, due to the intrinsic accessibility of the in-phase mode of coupling (C and V have been shown to be coupled in-phase in CV and anti-phase in VC; Goldstein, Byrd, & Saltzman, 2006). Linguistic universals are generally theory dependent and difficult to evaluate

(e.g., Newmeyer, 2005). The internal structural constraints of linguistic systems could be the cause (e.g., Hayes, 2004), but the origin could be biomechanical, either as a pattern unaffected by linguistic organization or as the source for the linguistic constraint itself. That is, if the source is biomechanical, there may be no need to posit a linguistic universal to explain the pattern.

Certain combinations of consonants and vowels have been discovered by MacNeilage and Davis to be preferred in both babbling and in dictionary counts of word types (Davis & MacNeilage, 1994, 1995; MacNeilage, 1998, 2008). Those authors observed that in early speech, labial consonants preferentially co-occur with central vowels, coronal consonants with front vowels, and dorsal consonants with back vowels. Rather than seeking an explanation in the markedness of the features, they proposed that the emergence of such preferred patterns is due to general physiological and anatomical principles. Also, they claimed that the neuromuscular activity related to chewing forms the basis of the preference for the CV combinations mentioned above. The alternation between mouth opening and closing that takes place during mastication was assumed to be phylogenetically and ontogenetically linked to the rhythmic oscillatory movement of the mandible. This oscillation, superimposed on phonation, generates cycles (Frames) that are perceived by the listener as syllables. During the early stages of language development, the infant exerts independent control only on the jaw, while other articulators (lips, tongue, and soft palate) have limited ability, if any, to actively vary their position in the brief span of a syllable. During jaw oscillation, if the tongue is in its resting position, the elevating movement of the mandible will make the lips form a passive constriction and produce a labial consonant, while jaw lowering will produce a central vowel; the authors assume that the tongue must be as distant from the alveolar ridge and the velar region as possible in order for the constriction to occur at the lips. Such a shape would result in a central vowel (MacNeilage & Davis, 1993, pp. 345–346). If the tongue is, by chance, in an advanced posture, jaw elevation will produce a coronal constriction since the front part of the tongue will make contact with the palate, while jaw lowering will produce a front vowel. In case of a (random) retraction of the tongue, jaw elevation will produce a velar constriction, and jaw lowering a back vowel (MacNeilage & Davis, 2000a, p. 287). The ability to exert control on articulators for the production of consonants and vowels (Content) is claimed to be acquired only in later stages of language acquisition, when motor control skills develop and the infant acquires a greater ability to master fine local movements of articulators within the time scale of a segment. The achievement of greater motor control allows inter-frame variation. According to MacNeilage and Davis, a first step from canonical to variegated babbling is represented by a particular intersyllabic pattern that seems to be present in infants' first words as well as in modern languages. It consists of a sequence made up of labial stop consonant – vowel – coronal stop consonant. The preference for this kind of pattern is called the Labial-Coronal effect (LC effect) and is considered to be the first step toward intersyllabic serial complexity (MacNeilage, Davis, Kinney, & Matyear, 1999, 2000). The LC pattern seems to be favored not only in babbling and early words, but also in adult speech (MacNeilage et al., 1999). This may be due to a preference for easier articulatory combinations even after content has been mastered.

A number of studies have also shown that the three co-occurrence patterns deriving from “frame dominance” are present in adult languages. Maddieson and Precoda (1992), while finding no overall evidence of “preference for articulatory convenience” (p. 55), do show “vowel deviation scores” that are consistent with the F/C predictions. More extensively, MacNeilage, Davis, Kinney, and Matyear (1999) analyzed data from dictionaries of 10 languages, namely, English, Estonian, French, German, Hebrew, Japanese, New Zealand Maori, Quichua, Spanish,

and Swahili. They reported that the ratios of observed to expected CV co-occurrence frequencies were greater than 1.0 only for the CV types favored in babbling and early words. Within those combinations, the labial-central combination is favored in seven languages out of 10, the coronal-front in seven languages out of 10, and the dorsal-back vowel in eight languages out of 10. Our own study of Mandarin dictionary counts showed partial support, with diagonals of 0.93, 0.94, and 1.62 (Whalen, Giulivi, Sam, Levitt, Hallé & Goldstein, submitted). The numbers for Mandarin are difficult to interpret given the affrication of coronals before high front vowels in the phonology. This not only removes a large number of potentially F/C compatible syllables, but it also changes the proportions in the other cells as well. With that in mind, the majority of reported adult language “type” counts show combinations that are in the preferred categories.

Dictionaries yield type counts, that is, frequencies of individual items regardless of their frequency of usage. Spoken corpora of adult speech are more similar to babbling by being “token” counts; that is, the frequency of occurrence of CV combinations is influenced by the relative frequency of particular lexical items. Babbling does not, as far as we can determine, contain types, so a token count is more comparable. Further, the token count is likely to reflect the infant’s experience of the language as well. In an analysis of two English corpora, two from French and one from Mandarin, Whalen et al. (submitted) found that the token counts gave different results for the three languages. English had two out of three diagonals greater than one, French had all three greater, and Mandarin had one out of three. Further, the English and Mandarin patterns were exactly opposite of each other. In an analysis of all nine cells (rather than just the diagonals), the English results were significantly *negatively* correlated with the babbling data, showing the universal tendency overshadowing the properties of the ambient language. The CV preferences posited by the F/C account thus can be modified in token counts by language-specific word frequency patterns, so any preference for it in babbling must be strong because that pattern is not consistently represented in the spoken input.

According to the F/C account, the relatively high frequency of the three CV patterns in adult languages, as well as in babbling and early speech, suggests that those patterns have been present since the origin of speech in hominids (Davis, MacNeilage, & Matyear, 2002; MacNeilage et al., 2000). This suggestion seems to be strengthened by a preference for the same CV patterns in a corpus of protolanguage words (MacNeilage & Davis, 2000b). The presence in adult languages of CV co-occurrences involving shared positions of the tongue, like the coronal-front and the velar-back combinations, is attributed to a basic common motor constraint favoring articulatory ease, namely, a constraint on the amount of lingual movement that needs to be made between the production of a consonant and the production of the following vowel (MacNeilage & Davis, 1993, 2000a). The presence in languages of the three basic patterns is also claimed to constitute for the infant “a ready-made initial access to the ambient language” (MacNeilage et al., 2000, p. 161). The English data in Whalen et al. (submitted) suggest that the pattern may provide access to language in general but not necessarily the ambient language.

The literature on mandibular control is divided on the question of whether speech and mastication use similar control mechanisms or not. Some studies find a great deal of similarity (Kent, Mitchell, & Sancier, 1991), some find small differences (Ostry & Flanagan, 1989), and others find fundamental differences (Moore & Ruark, 1996; Moore, Smith, & Ringel, 1988; Ostry, Vatikiotis-Bateson, & Gribble, 1997). In a recent study by Steeve, Moore, Green, Reilly, and McMurtrey (2008), an electromyographic analysis of the oromandibular activity during sucking, chewing, and babbling was conducted on 15 9-month-old normally developing children. The

results were compared to those obtained from 12- and 15-month-olds and from adults. It was observed that in 9-month-old children, the coordinative patterns observed during sucking were the same as in older toddlers, and the patterns observed for chewing and babbling were similar to those observed in 15-month-olds and in adults (with a progressive refinement, rescaling, and increase in stability with age). A comparison across behaviors showed that distinct, identifiable coordinative patterns were observable, respectively, for sucking, chewing, and babbling. The authors take this result as suggesting that the motor organization at the basis of babbling and speech may not be related to the coordinative organization of nutrition-related tasks, such as sucking or chewing. Given that a lack of difference is often difficult to interpret, the positive evidence of differences in organization must be taken seriously.

Articulatory Phonology (AP) (Goldstein, Byrd, & Saltzman, 2006) provides an alternative but still physiologically motivated explanation for the preference for certain kinds of CV associations in early speech. That is, in AP, the physiological basis is posited to be in the interaction of the articulators during babbling, not due to jaw motion alone (as in the F/C account). According to AP, the activity of the vocal tract can be analyzed in terms of “articulatory gestures,” namely, actions of constriction and release, by the different vocal organs, organized into temporally overlapping structures (Browman & Goldstein, 1989; Goldstein & Fowler, 2003). Individual gestures can be coordinated, or “phased,” by means of planning oscillators associated with the gestures that are coupled in intrinsically stable ways, corresponding to the stable modes exemplified by many simple dynamical systems, and in control and coordination of rhythmic movement (Haken, Kelso, & Bunz, 1985; Turvey, 1990). Planning oscillators associated with consonant gestures and vowel gestures can be coupled either in-phase or anti-phase. In CV syllables, the two gestures are hypothesized to be coupled in-phase (the most stable coupling mode). Since gestures are assumed to be triggered at a fixed phase of their associated oscillators, this means they are triggered synchronously and begin to control movements of the vocal tract articulators at the same time within the limits of articulatory compatibility. The gestures of two consecutive consonants, by contrast, are coupled in anti-phase mode and so are triggered sequentially. If the onset consonant cluster of, say, “spoon,” were produced synchronously, the tongue tip gesture for /s/ would be hidden by the lip gesture for /p/ and would not be perceptually recoverable. Instead, as Goldstein et al. (2006) point out, “a minimal requirement for a gestural molecule to emerge as a stable, shared structure in a speech community is that the gestures produced are generally recoverable by other members of the community” (p. 232). Consonant gestures and vowel gestures can be produced synchronously and still be recoverable due to the intrinsically different properties that characterize them. Vowel gestures are less constricted than consonant gestures, so that a consonant and a vowel gesture can be produced at the same time without the vowel interfering with the acoustic properties of the consonant. Furthermore, vowel gestures last longer than consonant gestures, so that in isolated CV syllables there is a period without overlap between consonant and vowel gestures.

Goldstein et al. (2006) have argued the favored CV combinations result from inherent articulatory compatibility so that coordinated movements to produce the C and V constrictions can begin synchronously, without any adjustments to their in-phase coupling. Articulations are compatible if they are either mechanically independent (as with the lips and tongue in /ba/) or constricted at the same locations (as with /di/ or /gu/). Nonpreferred CV combinations involve articulatory incompatibilities between the C and V so that their movements cannot begin completely synchronously without mutual interference (as with /du/ or /gi). The Articulatory Phonology

account is thus also based in the biomechanics of speech, but it makes different assumptions about the relevant control structures.

Both the F/C and the AP accounts of the preferred patterns—indeed, the existence of the pattern itself—depend on the division of the babbled utterances into regions of the vocal tract that are later to be related to linguistic features. Although it is not necessary to assume that the infant has adult-like categories in order to apply these labels in transcription, it is necessary to acknowledge that the transcription is an imposition and not a reflection of the infant's intention. It is likely that intention only becomes relevant when the infant begins using an articulatory routine repeatedly, but the nature of the earlier productions remains an interesting issue (Oller, 2000). There is enough agreement in transcriptions to give substance to the analyses, and enough of a relationship to biomechanics and to the ambient language to lend credence to the transcriptions. We can expect that new techniques for measuring speech production, such as ultrasound, will provide further refinement of our understanding of the substance of babbling. Also, brain imaging techniques such as near-infra-red spectroscopy (NIRS) may even provide evidence about intent. However, transcriptions continue to provide our best current evidence of the development of infant speech. All results based on transcription, such as our own and MacNeilage and Davis's, are subject to numerous pitfalls, such as native language bias and the bias toward hearing consonants and vowels in the first place. The current results must be placed within those limitations.

In sum, preference for certain CV co-occurrence patterns appears in babbling, first words, and, to a lesser extent, adult language. The F/C account posits that the early frame dominance in speech production gives way to greater articulatory control, as the infant "escapes frame dominance" (Davis & MacNeilage, 1995, p. 1208). At later stages, content elements are claimed to emerge in the form of actively independent movements, allowing the production of different vowels and consonants. According to AP, certain CV combinations are preferred (more frequent) both in babbling and adult speech as a consequence of the greater level of synergy (and consequent greater ease in syllable articulation) between the articulator gestures involved in the production of those combinations.

If the preference for certain CV syllables were to be attributed only to babblers' initial lack of articulatory control (as in the F/C perspective), then a decrease in the relative frequencies of those combinations should be expected with age, as the infant gains articulatory control.

In order to test whether successive stages of language acquisition are accompanied by a developmental trend in either direction in CV frequency, babbling data from three different ages were analyzed. According to the present view, as babblers mature, they gain more control over their articulators. Such increases in control should, on the F/C approach, lead to reduced (more adult-like) preponderance of the preferred CV combinations. We further explored this question by modeling the F/C and AP accounts via articulatory synthesis.

SEARCHING FOR DEVELOPMENTAL TRENDS IN THE FREQUENCIES OF CV ASSOCIATIONS

Method

We analyzed data from the Haskins babbling database (Levitt & Wang, 1991; Whalen, Levitt, Hsiao, & Smorodinsky, 1995; Whalen et al., 1991), containing longitudinal recordings and transcriptions of 17 babblers from three different linguistic backgrounds: French (six babblers),

English (six babblers), and Mandarin (five babblers). Children, all normally developing, were recorded biweekly between the ages of 5–6 months and 9–13 months. The transcriptions were done by a native speaker of Mandarin, who was also a speaker of English and Russian.

The procedure followed by Davis and MacNeilage (1995) was replicated on our data in order to test the F/C hypothesis. In particular, three categories of consonants and vowels were taken into consideration. Consonants included only stops, nasals, and glides (i.e., excluding fricatives) and were grouped, according to place of articulation, into labials (p, b, m, w), coronals (t, d, n, j), and velars (k, g, ŋ). We note that /w/ has, in adult language, both a labial and a velar component, but we follow MacNeilage and Davis in counting it only as a labial; its frequency is low enough that it would not change the results, had we instead (or additionally) counted it as velar. Vowels were grouped with reference to the front-back dimension of the vocal tract, into central (ʌ, ə, a, ø, ɜ, ʉ, i, ɜ), front (i, ɪ, e, æ, œ, y, œ, ø) and back (u, ʊ, ʉ, o, ɔ, ɑ, ɒ, ɐ, ɤ).

We searched the database for all the CV combinations whose consonants and vowels were among the sounds listed above. Given three consonant categories and three vowel categories, we searched, on the whole, for nine different types of CV combinations, namely, labial-central (B-A), labial-front (B-I), labial-back (B-U), coronal-central (D-A), coronal front (D-I), coronal-back (D-U), velar-central (G-A), velar-front (G-I), and velar-back (G-U). Henceforth, we will use capital 'B', 'D' and 'G', respectively, for labial, coronal, and velar consonants; and capital 'A', 'I', 'U', respectively, for central, front, and back vowels. Note that many of the sounds included, especially the vowels, are not found in the inventory of one or more of the target languages.

During the selection process, syllable position in an utterance was not taken into consideration, and diacritic marks and syllable boundaries (if marked in the transcriptions) were ignored. Segments connected by a ligature were categorized as the first element, and the second was ignored; for example, if a syllable was transcribed as [bə_æ] it would be categorized as labial-central.

Utterances that could be classified as words were not excluded from the analysis. However, the number of such utterances in the database was insignificant. During data collection, parents were explicitly asked to avoid any interaction with their children; therefore, no word elicitation took place in the sessions. Parent reports at the end of the study period for each infant yielded a varying number of words that the parents felt the infant had mastered, but most of the infants had no words reportedly under control.

As a second step, ratios of observed to expected frequencies for the nine CV combinations were generated for each babbler at each age. Next, in order to investigate whether the frequencies of B-A, D-I, and G-U combinations would show any developmental trend in the course of language development, we compared:

- i. the ratios of observed to expected frequencies for the favored CV combinations, generated from the English, French, and Mandarin babbling data from 6-month-old infants, with those generated from the data of the infants at 9 and 12 months of age. An inventory of the sounds produced by each babbler at each age was also made, in order to find out whether the variety of phones would increase over time, as would be expected if articulatory control were becoming greater.
- ii. the ratios generated for the three babbling ages with ratios generated from dictionary data, as representatives of each target language. In particular, for French and English we referred to dictionary data collected by MacNeilage et al. (1999), and for

Mandarin we analyzed data from an on-line Chinese-English Dictionary (CEDICT) (www.mandarintools.com/cedict.html). We are aware that a comparison between babbling and dictionary data is between data that are different in nature: Babbling data are tokens, whereas dictionary data are types; a separate study discusses the differences in the ratios based on token counts (Whalen et al., submitted).

- iii. the English ratios generated from the Haskins babbling database to ratios generated for babbling and first words by the authors of the F/C account. Specifically, we make reference to Davis and MacNeilage (1995) for babbling, and to Davis, MacNeilage, and Matyear (2002) for first words.

Transcription Reliability

The use of phonetic transcription in the study of infant vocalizations has often proved problematic. The high variability in the degree of well-formedness characterizing infants' sounds makes transcription intrinsically complicated and challenges inter-transcriber agreement. The need to describe infant sound productions, however, has led to developing different systems of sound classification, as well as new tools aimed at assessing transcription reliability.

Oller and Ramsdell (2006) propose a measure of transcription agreement that takes into account the different "weights" of phonetic differences between segments, on the basis of accepted principles of phonological theory. They analyze transcriptions of 30 infant utterances. After transcription pairing and alignment, they measure the proportion of slots for which transcribers agree on the presence of a segment and the proportion of phonetic features that are present in the transcriptions of segments belonging to the same slot. They calculate the overall agreement of transcriptions by multiplying the above two measures. In a 0 to 1 scale, they obtain average agreement of .60, whereas with a traditional "unweighted" measure, they obtain agreement of about .40.

Ramsdell, Oller, & Ethington (2007) stress the importance of evaluating "canonicity" of infants' productions and transcriber's confidence, as factors that can be predictive of transcription agreement levels. Syllables that are deemed by transcribers to represent firm control over the articulation (meeting "infrastructural requirements of well-formedness"; p. 794) are canonical and thus amenable to transcription. Precanonical utterances, on the other hand, are deemed too unstable to justify transcription: Agreement levels are low, as are transcriber confidence ratings. Noncanonical CV transitions are judged auditorily to be too slow to constitute adult-like consonant-vowel transitions (Oller, 2000). However, agreement levels on canonicity are also low (Ramsdell et al., 2007, p. 828). In our study, we assume that a consonant transcription is sufficient to place our stimuli in the canonical category, and thus amenable to transcription.

Koopmans-van Beinum, Clement, & Dikkenberg-Pot (2001) propose a transcription system (AMSTIVOC) that analyzes separately the phonatory and articulatory aspects of infants' productions. They identify five different types of phonation (according to whether phonation is continuous or interrupted) and three different types of articulation (depending on the number of articulatory movements enacted within one respiratory cycle). For consonants, they consider three places of articulations (front, central, and back) and five manners (fricative and trill; stop, glide, nasal, and lateral). Vowels are analyzed with respect to the position of tongue body (front, central, and back) and the degree of jaw opening (close, close-mid, open-mid, and open).

According to the authors, the system allows detailed investigation of infant vocalizations and permits the effective identification of differences in the productions of deaf and hearing children.

Stockman, Woods, & Tishman (1981) analyze transcription agreement in the productions of four children recorded from 7–21 months and coded by three different transcribers. They calculate agreement on identical segment matching and on-stop feature matching. They find that inter-transcriber agreement in the identical segment analysis ranges from 13–61%. In the stop feature analysis, they observe an overall mean agreement increase of 23%. Furthermore, the authors observe that, in general, agreement tends to increase with a child's age.

To investigate transcription reliability in the present study, 10% of the transcribed utterances of the Haskins babbling database were retranscribed by two listeners (one was an adult native speaker of American English and the other was of both Cantonese and American English), both trained in phonetic transcription but unaware either of the language environment of the babblers or of the preferred CV combinations. In line with the purposes of the present analysis, we calculated agreement for those CV syllables in which the consonants were stops, nasals, or glides. Agreement between transcribers was calculated on the basis of place of articulation matching. Consonants were classified as labial, dental, and alveolar and vowels as front, central, and back. We calculated agreement for consonants and vowels separately, but we also looked at whole syllable matching. We obtained 63% of agreement for consonants and 62% for vowels. The rates were less than those found by MacNeilage and Davis for transcriptions in which the transcribers compared their results and attempted to resolve discrepancies. This rate of agreement is similar to that found by Stockman et al. (1981) for babblers of the same age and with independent transcriptions.¹

Results

Our first step was to count the number of CV co-occurrences present in the Haskins babbling database across the three categories of consonants and vowels detailed earlier. The counts for each of the three ages (6, 9, and 12 months) are given in Table 1.

From the counts reported in Table 1, ratios of observed to expected frequencies for the nine CV combinations were generated (see Table 2). The numbers in bold face in the diagonals correspond to the ratios of observed to expected frequencies for the three types of CV combinations predicted as favored by the F/C account, namely labial-central, coronal-front and velar-back.

¹Further validation of transcriptions was provided through a two alternative forced choice identification test, conducted to determine whether listeners from two language backgrounds would agree not only with the transcriptions used in the present study, but also with acoustic indications of consonant and vowel identity. The aim was to understand whether there is an intrinsic bias towards hearing the favored CV combinations or not. Participants were two groups of native American-English speakers and two groups of native Italian speakers, who were asked to listen to four different kinds of CV syllables (labial-central, coronal-front, labial-front and labial-central), and judge the place of articulation of consonants and vowels.

A repeated measures ANOVA was performed on the results of both the consonant and the vowel test. The between factor was Language (English or Italian) and the within factors were Consonant and Vowel. The dependent variable was the percent of responses that agreed with the transcriber's judgment.

Overall the test showed good agreement both with the transcriptions considered for the present study and with the acoustic indications of sound identity (even if vowels were not as accurately identified as the consonants). No noticeable tendency towards hearing the preferred CV combinations was observed.

TABLE 1
Number of Occurrences of Consonant and Vowel Places of Articulation Produced at 6, 9 and 12 Months by the French-, English-, and Mandarin-Learning Babblers

<i>Cons</i>	<i>6 Months</i>				<i>9 Months</i>				<i>12 Months</i>				<i>Grand Total</i>
	<i>B</i>	<i>D</i>	<i>G</i>	<i>Tot</i>	<i>B</i>	<i>D</i>	<i>G</i>	<i>Tot</i>	<i>B</i>	<i>D</i>	<i>G</i>	<i>Tot</i>	
FRENCH													
Vow													
A	76	34	25	135	20	72	11	103	66	142	24	232	470
I	117	147	70	334	55	218	61	334	54	279	98	431	1099
U	28	14	12	54	3	26	5	34	11	17	10	38	126
Grand total	221	195	107	523	78	316	77	471	131	438	132	701	1695
ENGLISH													
Vow													
A	18	76	45	139	82	147	87	316	114	263	142	519	974
I	93	533	255	881	202	621	254	1077	165	539	316	1020	2978
U	7	22	15	44	25	31	31	87	24	23	34	80	211
Grand total	118	631	315	1064	309	799	372	1480	303	825	492	1619	4163
MANDARIN													
Vow													
A	23	12	17	52	37	11	15	63	76	118	85	319	394
I	42	95	68	205	36	135	91	262	151	913	475	1539	2006
U	1	5	2	8	20	2	2	24	7	76	16	99	131
Grand total	66	112	87	265	93	148	108	349	234	1107	576	1917	2531

TABLE 2
Ratios of Observed to Expected Frequencies Obtained From the Data of the Haskins Babbling Database

<i>Cons</i>	<i>6 Months</i>			<i>9 Months</i>			<i>12 Months</i>		
	<i>B</i>	<i>D</i>	<i>G</i>	<i>B</i>	<i>D</i>	<i>G</i>	<i>B</i>	<i>D</i>	<i>G</i>
FRENCH									
Vow									
A	1.33	0.68	0.91	1.17	1.04	0.65	1.52	0.98	0.55
I	0.83	1.18	1.02	0.99	0.97	1.12	0.67	1.04	1.21
U	1.23	0.70	1.09	0.53	1.14	0.90	1.55	0.72	1.40
ENGLISH									
Vow									
A	1.17	0.92	1.09	1.24	0.86	1.09	1.17	0.99	0.90
I	0.95	1.02	0.98	0.90	1.07	0.94	0.86	1.04	1.02
U	1.43	0.84	1.15	1.38	0.66	1.42	1.60	0.56	1.40
MANDARIN									
Vow									
A	1.78	0.55	1.00	2.20	0.41	0.77	2.23	0.73	1.01
I	0.82	1.10	1.01	0.52	1.22	1.12	0.80	1.03	1.03
U	0.50	1.48	0.76	3.13	0.20	0.27	0.58	1.33	0.54

In the French data, at 6 months, ratios along the diagonal are greater than 1.0 for all the favored CV combinations. At 9 months only one on-diagonal ratio exceeds 1.0, whereas at 12 months they increase, with labial-central and velar-back combinations reaching values greater than those recorded at 6 months.

For English, the three combinations predicted as favored show ratios >1.0 at all ages. Values record an increase from 6 to 9 months and then a decrease from 9 to 12 months.

For the Mandarin data, only the labial-central and the coronal-front co-occurrences are greater than 1.0. In general, the data show particularly high ratios for the labial-central combination and particularly low ratios for the velar-back.

In general, as the results show, the babbling of the three languages presents greater than 1.0 ratios of observed to expected frequencies for the three CV combinations predicted as favored by the F/C account. However, there are CV combinations, other than the favorite ones (off-diagonals in the table), that also present relatively high frequencies. According to AP, this is to be interpreted as being from some articulatorily synergetic factor that is at work both for the preferred and nonpreferred syllables since the very early stages of language development. If, as in the F/C account, the preference for the diagonal combinations were to be attributed only to jaw oscillation (lack of control on other articulators), we would expect to see much higher ratios in the diagonal cells and much lower ratios in the off-diagonals. Instead, as the table shows, coronal consonants combine frequently with back vowels in the French babbling at 9 months and in Mandarin babbling at 6 and 12 months. A similar result is found for French in a study conducted by Boysson-Bardies (1993) on CV associations in disyllabic productions of 10- to 12-month-old learners of French, American, Swedish, and Yoruba. In the same study, high frequency of labial-back combinations is found for Swedish (even if only in second syllable position). In our data, labial consonants combine frequently with back vowels in the babbles of 6 and 12 months French infants, in those of 6, 9 and 12, months English and in the productions of 9 months Mandarin. The labial-back combination has been found to be preferred in Mandarin also in a study by Chen and Kent (2005), where the authors analyze the productions of Mandarin-learning infants, recorded from 7–18 months.

Velar consonants combine with front vowels more often than expected in our French and Mandarin data at all ages and in the English data at 12 months. High frequency of velar-front associations is found in the Yoruba data of Boysson-Bardies (1993). In the same study, the American and Swedish data present high frequency of velar-central combinations. A similar result is found in our data for English (at 6 and 9 months) and Mandarin (at 6 and 12 months) even if the values are generally smaller.

Testing for Developmental Trends

Our first statistical test replicates that of Davis and MacNeilage (1995), a chi square test, in order to be as comparable as possible. Combining the three languages in tests at each age, the test showed that the differences between the observed and the expected frequencies for the predicted CV combinations were significant at all ages (6, 9, and 12 months), given that all consonants can combine at random with all vowels. Seven out of nine predictions (three predictions for each of the three age stages) were confirmed. The two cases in which the expected preference did not occur involved the velar-back combination at 6 and 12 months.

A more stringent test was run in which babblers were taken as individual cases in a repeated measures Analysis of Variance (ANOVA), with the within factors of Syllable Type (BA, DI, GU) and Age (6, 9, and 12) and the between factor of Language (English, French, and Mandarin) for the dependent measure of the ratio. The two missing values at the 12-month level were replaced with the overall mean for the remaining 12-month data. The one missing value at 9 months was similarly replaced with the mean for all 9-month data. Ratios were decreased by 1, so that the test of the overall mean would indicate whether the ratios were significantly greater than 1 or not.

Results for the three main factors and all interactions were nonsignificant, with *F* values ranging from 0.39 to 2.08 (all *p*'s > .05). The overall mean, however, was significantly greater than 0 ($F(1,14) = 5.59, p < .05$), with an average value of 1.24. The overall tendency to have diagonals greater than 1 was supported, which is in line with both F/C and AP reasoning. The lack of a difference across language environments is also consistent with both accounts. However, the lack of any indication of change by age is inconsistent with the F/C account but consistent with the AP account.

To investigate whether it was possible to observe a developmental trend in the frequencies of the CV combinations that are claimed to be favored as a consequence of frame dominance, we plotted in Figure 1 the ratios of observed to expected frequencies obtained for each language at each age.

As already mentioned, the F/C account claims that, as language develops, the role of jaw oscillation ('frame') in shaping speech production becomes less dominant and children acquire greater control over other articulators. Therefore, we expected to observe, between 6 and 12 months, a developmental decrease in the frequencies of the predicted combinations. Instead, as Figure 1 shows, no consistent developmental trend was observable in the ratios of any of the three languages. This would imply that jaw oscillation is still playing a major role in the children's productions and that no gains in articulatory control took place in the time span of six months. However, a count of the different phones produced by each babbler at each age seems to contradict this conclusion. Table 3 reports the number of different phones produced by each infant at 6, 9, and 12 months. The average number of different phones produced by infant learners of each language at each age is also given at the bottom of Table 3.

Only five of the six French participants were included in the count. The one we left out (JZ) had been recorded at 6 months only. The two infants who were recorded at 6 and 9 months, but not at 12 (MB and FL), were retained.

Table 3 shows that the most common tendency is for the number of sounds produced by each babbler to increase progressively from 6 to 12 months. This happens for example in the babbles of infants MA, MM, and NG for English; MB and YC for French; and EW, FL, TZ, and YL for Mandarin. In a few cases, the number of phones decreases at 9 months and increases again at 12 months. Looking at the transcriptions, we noticed that this pattern is observed when the number of utterances at 9 months is very limited. Therefore, the reduction of the number of phones can be attributed to the limited amount of data. This is the case for babblers AB, EC, and BX, respectively, for English, French, and Mandarin. In some other cases, the number of different phones tends to diminish from 6 to 9 and from 9 to 12 months, for example, for the English babbler VB and the French babbler MS. Again, looking at the transcriptions we noticed that this happens only when the inventory of sounds is already very large at 6 months, so that the 9- and 12-month counts may under-represent the true inventory. Finally, in English CR and

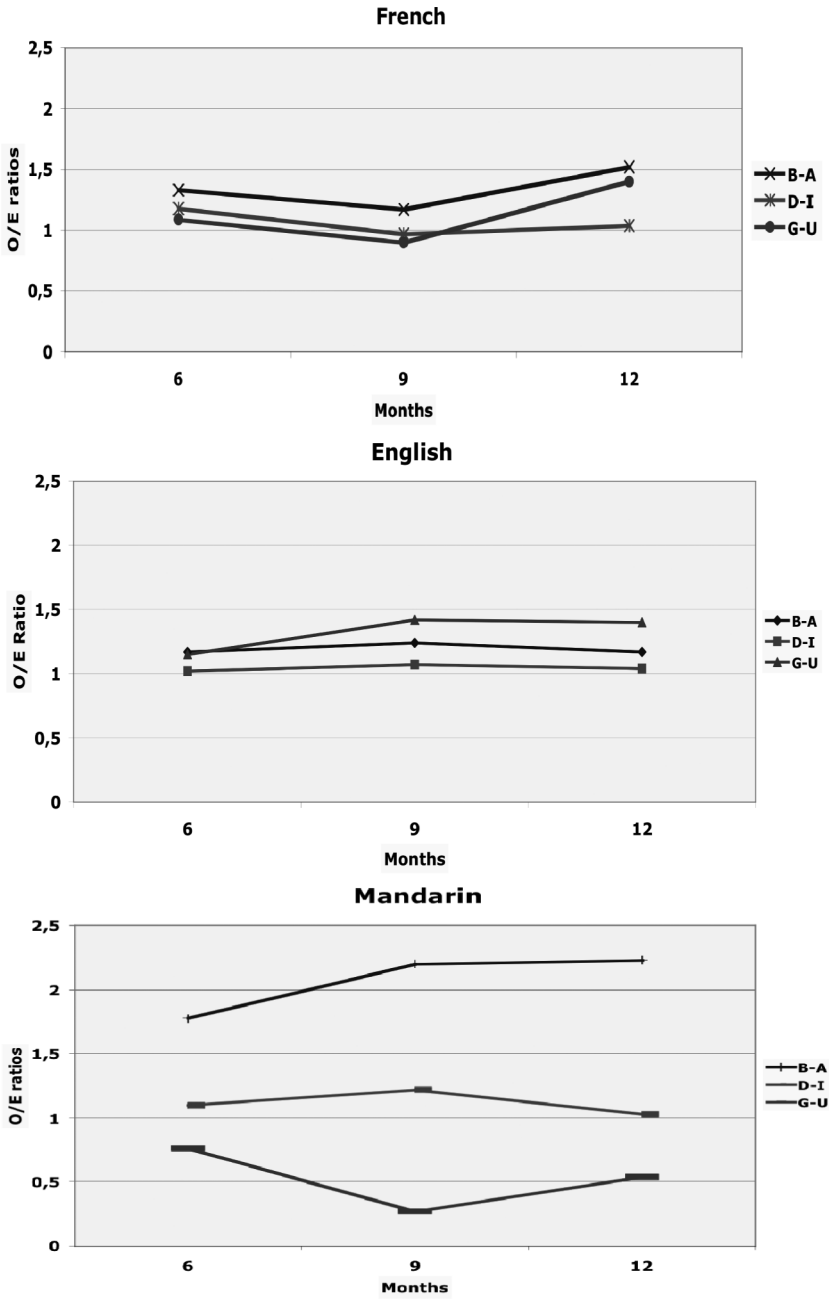


FIGURE 1 Ratios of observed to expected frequencies of the three favored CV co-occurrences, for the French-, English- and Mandarin-learning infants.

TABLE 3
Number of Different Phones Produced by Each Infant in the Haskins Database
at 6, 9 and 12 Months of Age

<i>Babbl.</i>	<i>Lang.</i>	<i>6 Months</i> <i>N° of Sounds</i>	<i>9 Months</i> <i>N° of Sounds</i>	<i>12 Months</i> <i>N° of Sounds</i>
AB	English	27	25	28
CR	English	15	27	23
MA	English	18	22	35
MM	English	20	23	27
NG	English	25	28	31
VB	English	36	34	31
EC	French	25	21	34
MB	French	20	21	
MS	French	37	35	34
NM	French	11	26	23
YC	French	25	29	36
BX	Mandarin	29	14	22
EW	Mandarin	15	16	29
FL	Mandarin	18	22	
TZ	Mandarin	25	27	35
YL	Mandarin	13	24	41
Average	English	23.5	26.5	29.1
	French	23.6	26.4	31.75
	Mandarin	20	20.6	31.75

French NM, the number of distinct sounds increases from 6 to 9 months and then decreases from 9 to 12 months. Overall, there is evidence of increased control over the speech articulators during this six-month period.

Comparison with Adult Language

In order to further probe the issue of the existence of a developmental trend, the babbling ratios of observed to expected frequencies generated from the Haskins babbling database were compared to those obtained from French, English, and Mandarin dictionary data. The aim was to see whether it was possible to observe any developmental trend, from babbling to adult languages, in the relative frequencies of the CV combinations predicted as favored by the F/C account. The French ratios for the B-A, D-I, and G-U syllable types were 1.2, 1.6, and 1.6, respectively. For English the ratios are 1.2, 1.0, and 1.5, and for Mandarin, 1.2, 1.2, and 1.9.

There is no consistent developmental decrease in the ratios between babbling and the dictionary values, that is, toward target language. Indeed, some of the dictionary values are larger than those of the babbling.

In Table 4 we report the Haskins babbling ratios generated for English and the average ratios obtained for babbling, first words, and dictionary, respectively, in Davis and MacNeilage (1995), Davis, MacNeilage, and Matyear (2002), and MacNeilage et al. (2000).

No systematic evidence of differences among babbling, first words, and adult languages is observable in the table. First word ratios are higher than the babbling ratios for two out of three

TABLE 4
 Ratios of Observed to Expected Frequencies of the Favored CV Combinations
 Generated for English Babbling, First Words and Dictionary Data

	<i>Age (months)</i>	<i>B-A</i>	<i>D-I</i>	<i>G-U</i>
ENGLISH				
Babbling (Haskins)	6	1.17	1.02	1.15
	9	1.24	1.07	1.42
	12	1.17	1.04	1.40
Babbling M&D	6.5–12	1.39	1.41	1.34
First words M&D	12–18	1.30	1.43	1.38
Dictionary		1.18	1.0	1.5

CV combinations, namely coronal-front and dorsal-back. The F/C account attributes this lack of change to the increased “functional load” that word learning involves (Davis et al., 2002). It is possible that the perceptual reorganization that accompanies the acquisition of first words (Werker, 1989) plays a role, but no direct mechanism has been proposed. Nonetheless, within the age range of babbling, no increase in functional load is evident, and so the F/C account should lead us to expect a progression toward adult values as articulatory control increases. According to AP, the favored CV combinations emerge in babbling as a consequence of the greater ease of articulation involved in their production and are eventually retained also in later stages of language development.

Discussion

The analysis described so far has focused on testing the co-occurrence of consonants and vowels in babbling for preferences according to place of articulation. Overall, as predicted by both the F/C and the AP account, the observed-to-expected ratios for these syllables were significantly greater than 1 (1.24), indicating that the preference was present in our data. Some details were not consistent with this overall effect: Only the English data show ratios greater than 1.0 for all three favored combinations at all ages, the French data give ratios less than 1.0 for the coronal-front and velar-back combinations at 9 months, and in the Mandarin data the velar-back combination is always less than 1.0 at all ages.

As already mentioned, according to the F/C account, the regularity of mandibular oscillation provides the basis for what the listener perceives as a syllable-like input. In canonical babbling, no other neuromuscular activity is said to be associated to the movement of the mandible, and therefore no subsyllabic organization (Content) should be present during the jaw cycle (Frame). As a consequence of this kind of organization, the account predicts the three favored CV co-occurrence patterns mentioned earlier. Such co-occurrences are claimed to be due only to the anatomical and acoustic consequence of jaw movement, with lack of lingual independence. There is no combination (of C and V) taking place. However, if babbling were really dominated by the movement of the mandible, without any independent action of other articulators, we would expect much higher ratios than those obtained both in our data and in those of MacNeilage and Davis. It is true that MacNeilage and Davis’ English babbling results give ratios

of observed to expected frequencies always greater than 1.0, and the same can be observed in our data (even if values are generally smaller). However, the average ratios reported in Davis and MacNeilage (1995) are less than 1.5; the underlying counts (provided by the authors) indicate that 51.6% of the productions fall into these preferred categories. The figure for the results in Davis, MacNeilage, and Matyear (2002) is 49.9%.

MacNeilage and Davis's data show ratios greater than 1.0 for many CV combinations other than the favored ones. This is true also for our French and Mandarin babbling data as well as for some Italian babbling data found in the literature (Zmarich & Miotti, 2003). Since, according to the account, jaw oscillation is responsible only for the three favored CV combinations, the high frequencies of the nonfavored ones must be attributed to something different than frame dominance. As we will see in the following paragraph, according to AP, the (slight) preference for the favored CV combinations is due to the more synergetic articulatory actions involved in the production of consonants and vowels.

Furthermore, the fact that the inventory of sounds produced by the majority of infants in the Haskins database tends to increase from 6 to 12 months seems to suggest that children are acquiring greater control over articulators and freeing themselves from the 'frame' bias. Indeed, some preliminary findings by Serkhane, Schwartz, Boë, Davis, & Matyear (2007) seem to suggest an increase in articulatory exploration already between 4 and 7 months of age.² The authors compare acoustic data of 4- to 7-month-old infants vocalizations to the productions of an articulatory-acoustic model of the growing vocal tract. The discrepancies between the data obtained from the simulations and the infants' data suggest that the latter could have been the result of greater complexity of movements rather than of jaw oscillation alone.

Indeed, if jaw oscillation were the only factor responsible for the preferred CV combinations, as in the F/C account, the achievement of greater articulatory control should lead to a developmental decrease in the frequencies of the three CV combinations predicted to be favored as a consequence of frame dominance. This means that smaller ratios of observed to expected frequencies should be observable for those combinations. The ratios generated from our babbling data do not show any such developmental trend. Nor is it possible to find a consistent decrease in the ratios reported by MacNeilage and Davis for babbling, first words, and adult data. The lack of frequency decrease is in line with the explanations offered by AP for the preferred CV combinations.

In general, for all of our adult data (both spoken language and dictionary data), as well as for those of MacNeilage and Davis, we obtained ratios of observed to expected frequencies greater or very close to 1.0. This means that there are relatively high frequencies of the CV combinations that are claimed to be a consequence of the sole oscillatory movement of the mandible also in adults. As mentioned in the introduction, MacNeilage and Davis explain this residue as a slight preference for greater ease of articulation.

As already mentioned, Articulatory Phonology offers a different explanation for these results, as will be described in the following paragraph.

²The acoustic results of Serkhane et al. (2007) show some correspondences with those of Sussman et al. (1999). In this study, the authors recorded a child from 7 to 40 months of age. They found that CV coarticulation, as indicated by locus equations, grew increasingly adult-like between 7 and 13 months.

MODELING OF FRAME THEN CONTENT AND ARTICULATORY PHONOLOGY DYNAMICS

Articulatory Phonology proposes an alternative explanation for the relatively high frequencies of these preferred CV combinations in children and adult speech (Goldstein et al., 2006). In CV syllables, the gesture of the onset consonant and the gesture of the following vowel are hypothesized to be initiated synchronously (in-phase). However, not all CV combinations can allow articulatory synchrony due to anatomical limits of the articulators involved and to the demands on shared articulators. For example, one of the CV combinations predicted as disfavored by the F/C account is a coronal consonant with a back vowel, for example, [du]. Producing the tongue tip gesture for [d] at the same time as the back vowel gesture for [u] may not be as easy to achieve as producing the same consonant gesture at the same time as the gesture for a front vowel such as [i] because in the coronal-back combination the tongue body needs to be fronted for the achievement of the tongue tip constriction, but the production of the back vowel requires the tongue body to be retracted. Therefore, the consonant and vowel gestures may not be produced synchronously. In other words, the vowel gesture may need to be delayed.

The three CV combinations predicted as favored by the F/C account, do not present such difficulties. In the coronal consonant-front vowel combination ([di]), for example, the tongue body needs to be fronted for producing both the tongue tip constriction and the front vowel. Therefore, the two gestures can be synchronized. The same can be said for the labial consonant-central vowel combination: During jaw raising, the tongue needs only to remain in its resting position, therefore no undesired constriction is produced. With the velar consonant-back vowel combination ([gu]), the tongue body needs to be retracted for the production of both the velar constriction and the back vowel, so that synchronization of the two gestures can easily be achieved.

Some recent evidence showing coordination differences between articulatorily compatible C-V sequences and incompatible ones has been found by Gao (2008). She measured tongue, lip, and jaw motion via electromagnetometry (Perkell, Cohen, Svirsky, Matthies, Garabieta, & Jackson, 1992) for a set of 36 syllables produced by seven speakers of Mandarin Chinese. Their onsets for C and V gestures in labial-central vowel sequences were more nearly synchronous than those in the coronal-central vowel sequences.

According to AP, certain CV combinations are preferred over others because they can be produced in-phase. This means that their production is characterized by the fact that articulatory synchrony matches synchrony in gestural triggering. For this reason, those combinations can emerge and be produced more easily, and this could explain their relatively high frequencies both in babbling and in adult languages. Goldstein et al. (2006) point out that adults may have maintained this preference for those CV combinations that can be synchronized. The oscillatory movement of the mandible could conceivably still play a role in favoring those combinations, but “the existence of these CV preferences cannot constitute evidence for a jaw-only motor control strategy, since the preferences exist in adult languages but the jaw-only strategy does not” (Goldstein et al., 2006, pp. 238–239).

The preferred combinations that are apparent in the data outlined by MacNeilage and Davis and in our own work seem amenable to a biomechanical explanation, but there is no set of articulatory data that can test the question. Not only is it difficult to study the articulation of infants, but it also is not completely clear even for adults what would constitute the correct line

of evidence showing biomechanical preferences. Here, we will present a summary of efforts to model both accounts.

The modeling makes use of a stylized but plausible articulatory synthesizer, the Configurable Articulatory SYnthesizer (CASy) (Iskarous, Goldstein, Whalen, Tiede, & Rubin, 2003) that can be set to perform kinematically determined movements or be driven by a task-dynamical model of speech (TADA) (Nam, Goldstein, Saltzman, & Byrd, 2004). Here, we will briefly describe how we used each mode to model the F/C account and the AP account. Other technical issues are described more fully in a companion paper (Nam, Giulivi, Goldstein, Levitt, & Whalen, submitted).

Frame Then Content Predictions for Jaw-Only Movement

As summarized earlier, the F/C model assumes that the supralaryngeal articulators begin at a random configuration that is then closed and opened by jaw motion alone (Davis & MacNeilage, 2002, p. 139). Because CASy allows changes in overall vocal tract length, we also modeled the default, adult setting, and a scaled version matched to the MRI images of a 7-month-old infant (Vorperian, Kent, & Lindstrom, 2005). The jaw is a separate articulator in CASy (uncorrelated with other articulators), so we were able to reduce movement to this one parameter. Tongue location was varied systematically such that the length and angle of the tongue body relative to the jaw, and of the tongue tip relative to the tongue body, were randomly selected within physiologically reasonable bounds. The jaw was then raised until some portion of the vocal tract was obstructed; this point of contact was classified as labial (lips made contact), alveolar (tongue tip made contact), or velar (tongue body made contact). The jaw was then lowered a randomly varying amount, simulating a range of vowel heights. Results were classified as front, central, and back for the vowels by comparing the shapes to MRI images for static vowels and finding the nearest match. Ratios were then computed as for the transcriptional data (see Table 5).

Diagonal ratios in the adult model were 1.01, 0.96, and 2.44 for labial/central, coronal/front, and velar/back, respectively. Diagonal ratios in the infant model were 0.86, 1.08, and 3.67 for labial/central, coronal/front, and velar/back, respectively. Two out of three ratios are above one, as predicted by the F/C account. However, the 1.01 and 1.08 ratios are barely above 1,

TABLE 5
Ratios Generated by the Simulation of the
Frame/Content Account. A) Adult Vocal Tract,
Varying Location of Consonant Closure
Location. B) Vocal Tract Scaled to a
7-Month-Old's MRI Images

	<i>B</i>	<i>D</i>	<i>G</i>
A.			
A	1.01	1.34	0.63
I	1.06	0.96	0.82
U	0.92	0	2.24
B.			
A	0.86	1.24	0.73
I	1.61	1.08	0.04
U	0	0.31	3.67

so that these models provide little support for a jaw-only source of the CV preferences. The off-diagonals are also variable, ranging from 0 to 1.61. Thus, there is no real support for the assumptions of the F/C account from either of the models run here.

Using Articulatory Phonology Synergies to Predict Preferred Co-occurrences

To test the AP account, we reasoned that if a CV combination is highly synergistic (and therefore affords synchronization of C and V gestures), the tongue movements exhibited by the synchronized activation of those two gestures (C and V gestures) will be similar to each other. The task-dynamic implementation of AP (TADA) was used to simulate this. We produced bare consonant and vowel trajectories and compared their final configurations of the tongue body between the gestures. The smaller the difference in the end positions between C and V gestures, the more synergistic the CV combination. The parameters that are needed for this model have never been computed for infant vocal tracts, so it was impossible to model them. The similarity of the infant and adult models for the F/C account indicates that it is reasonable to expect similar results if the TADA model were to incorporate such radically different vocal tracts.

To generate ratios from these synergies, we first found the magnitude of the difference of the path for the C and for the V; smaller differences are equivalent to greater synergy. Eleven values were used for the C: a single value for labial and five equally spaced locations along the palate for each of alveolar and velar place. The vowel space was sampled by distributing 300 target values across all physiologically possible locations of the center of the tongue. The values for all 3,300 CV combinations (300×11) were then normalized to range from 0 to 1; this value was then subtracted from 1 to give a likelihood estimate (given that small values are more likely). The sum of these likelihoods for each of the nine CV cells was then computed, with the ratios of obtained to expected calculated over the total sum.

The results are shown in Table 6. All the diagonals have the highest values, as found in the transcription data. The values for the off-diagonals were also well predicted.

Comparison of the Two Models

The F/C models, adult and infant, generated only two out of three diagonals greater than one, while the AP model predicted all three. These models were compared with the off-diagonals

TABLE 6
Ratios from Relative Likelihood of
Co-occurrence of CV Places of
Articulation from the Articulatory
Synthesis Implementation of Articulatory
Phonology

	<i>B</i>	<i>D</i>	<i>G</i>
A	1.29	0.79	1.14
I	0.73	1.32	0.74
U	1.20	0.45	1.50

as well, even though the original explanation for the F/C account did not encompass the off-diagonals. The magnitude of the differences between the transcriptional data and the model is much smaller for the AP model. For the three diagonals, the average difference between the model and the data is 0.15 for the AP account and 1.09 for the F/C. For the six off-diagonals, the values are 0.23 and 0.71. Thus even though the original explanation did not make any predictions for the off-diagonals, those are even more accurate than the diagonals. The error for the AP account is less than a seventh of the F/C account. Therefore, while both accounts provide a physiological rationale for the preferred combinations, the AP account provides a better fit with the data.

Two other models have also failed to show substantial support for the F/C account. Serkhane et al. (2007) used a different articulatory synthesizer (Maeda, 1990) and found that some of the predicted patterns were confirmed, but some categories were greatly different from predictions. Even this limited success must be viewed with some caution, given that the “jaw” component in the Maeda model incorporates some elements of tongue motion since it is derived from Principal Components Analysis of x-rays of adult speech (see Nam et al., submitted, for further details). A second approach is found in Lindblom (2008), who calculated degree of compatibility as amount of difference between the tongue shapes for consonants and for vowels. He found that alveolars were most compatible with front vowels, but velars were most compatible with front/central vowels and became increasingly incompatible with back vowels. Thus neither other modeling approach supports the F/C account.

Several features of the AP model differ from the F/C. First, the consonant constriction is taken as a primary goal. Whether or not infants have a particular goal when they babble is currently not addressable. Nonetheless, the (variably placed) consonant constriction is assumed to be controlled in this account. Similarly, the vowel is not assumed to be fully determined by jaw opening but by some combination of jaw opening and a further specification of a tongue target. As previously noted, the F/C account would generally predict that **all** utterances would be on the diagonal in the place of articulation tables. That is, the reason the F/C account predicts the preference for the diagonal cells is that random starting positions will primarily result in the predicted CV types. However, none of the models of the jaw-only control of articulation results in the patterns found in babbling (Vilain et al., 1999; Serkhane et al., 2007; Nam et al., submitted). In the Nam et al. study, the F/C simulation did result in about half of the syllables being on the diagonal (as in the babbling data), but there was no correlation with the ratios found in the babbling data. It is therefore unclear what the off-diagonal combinations entail for the F/C account and whether the original rationale of the model can be said to be viable. For the AP account, there is relatively good correspondence between the ratios both on and off the diagonals (Nam et al., submitted). Articulatory synergy, therefore, seems to be greatest on the diagonals, as originally observed by MacNeilage and Davis, but is also a factor in the off-diagonals.

Furthermore, it also accounts for another frequently observed characteristic of babbling: a (near-)universal preference for CV syllables over VC. To our knowledge, the F/C account does not provide a ready explanation for this effect. If the articulators are carried by the jaw, it would seem that just as many accidental closures would appear to be VC as CV. If, on the other hand, the infants are combining C and V actions that they (at least partially) control, they would be expected to employ only the simplest coupling mode (in-phase), and only CV syllables would result. The consonant of VC syllables is in an anti-phase relationship to the vowel, so VC syllables are intrinsically more difficult to produce and would be expected to develop

later. Nam, Goldstein, and Saltzman (2009) show in simulations of infant development that the earlier production of CV syllable can result from the greater intrinsic accessibility and strength of in-phase coupling.

GENERAL DISCUSSION

There is a pattern of favored co-occurrences of consonants and vowels in babbling and, to a certain extent, in adult language (Davis & MacNeilage, 1995). We explored an implication of the F/C account, that is, that the proportion of such favored pairing should decrease as articulatory control increases. Data from the babbling of children from French, English, and Mandarin language environments were used. As an overall replication of the main result, the favored patterns occurred more frequently than chance would predict, for the most part, as seen in ratios greater than one. There were some instances for the French and Mandarin data for which these ratios were less than 1, contradicting the favored pattern. For all of the cases, the magnitude of the ratios was usually small enough that it represented only a plurality of such CV combinations in the data. By direct measurement, these combinations accounted for approximately 50% of the total in both our data and that of MacNeilage and Davis.

There was no developmental trend in the ratios of observed to expected frequencies of the CV combinations predicted to be favored. The F/C account leads to an expectation of a progressive decrease over development in the frequencies of the syllable types predicted as favored as a consequence of frame dominance. This should occur because the account claims that the effect of jaw oscillation becomes smaller as articulatory control increases. The Haskins babbling database covers a time span of six months with the utterances we analyzed having been produced at 6, 9 and 12 months of age. The data in the database did not exhibit any consistent developmental decrease in the frequencies of the favored CV combinations, nor did a comparison between the babbling data and target language data, coming from French, English, and Mandarin dictionaries.

Articulatory Phonology offers an alternative explanation: the relatively high frequencies of the three CV combinations are due to the relative ease with which independent (BA) or nearly identical (DI and GU) gestures can easily be produced synchronously. This happens for example in CV syllables made up of a coronal consonant and a front vowel, where both the consonant and the vowel gesture involve an advancement of tongue body. In other combinations, the consonant gesture and the vowel gesture may involve motions of the same articulator in opposite directions and may therefore be competing to reach their respective targets. For example, syllables made up of a coronal consonant and a back vowel will need an advancement and a retraction of tongue body, respectively, for the production of the consonant and of the vowel.

Matyear (2007) suggested that there are shallower transitions for babbled utterances than for adult speech. Indeed, for the babbled utterances, there was virtually no evidence of formant transitions at all. This would, in fact, show a lack of control over the articulators if the measurements are accurate, but there are two reasons for thinking they are not. First, the high fundamental frequency of the voice source in infant speech makes it difficult to track anything other than the most intense harmonic of a formant. Thus even if the formant is moving, it will look as if it is not, since only one harmonic is likely to be measured. The other consideration is that transcribers do, in fact, hear consonants at different places of articulation (fully 50% of the “diagonals” in

both our data and that of MacNeilage and Davis). If there were no formant transitions on which to base this judgment, then we would be hard pressed to explain this consistent transcription. Therefore, it seems unlikely that these tantalizing results will hold up. The acoustic evidence, then, is more in line with a lack of change during maturation.

The slightly greater than expected co-occurrence of the “favored” combinations does, as MacNeilage and Davis originally pointed out, call for a physiological explanation. Given the lack of evidence for a developmental trend, the persistence of the pattern in adult type counts, and the inconsistency of the pattern across languages in adult speech, it appears that the pattern is not due to cyclicity of the jaw. Instead, the greater ease with which some consonant and vowel gestures combine with each other, as predicted by Articulatory Phonology, explains both the existence the recurrence in different circumstances, and the variable influence in adult speech of CV combination preferences.

ACKNOWLEDGMENTS

This work was supported by NIH grants DC-000403 and DC-002717 to Haskins Laboratories. Portions of this work appeared in an unpublished Ph.D. dissertation (Giulivi, 2007). We thank Carol A. Fowler, Michael Studdert-Kennedy, Julia Irwin, Aude Noiray, D. Kimbrough Oller, and two anonymous reviewers for helpful comments. Data for calculating percentages of preferred response from previous studies were kindly provided by Barbara L. Davis and Peter MacNeilage.

REFERENCES

- Boysson-Bardies, B. de (1993). Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 353–363). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Boysson-Bardies, B. de, Hallé, P. A., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, *16*, 1–17.
- Boysson-Bardies, B. de, Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target-language. *Journal of Child Language*, *11*, 1–15.
- Boysson-Bardies, B. de, & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and early words in four languages. *Language*, *61*, 297–319.
- Breen, G., & Pensalfini, R. (1999). Arrernte: A language with no syllable onsets. *Linguistic Inquiry*, *30*(1).
- Browman, C. P., & Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 151–206.
- Chen, L.-M., & Kent, R. D. (2005). Consonant-vowel co-occurrence patterns in Mandarin-learning infants. *Journal of Child Language*, *32*, 507–534.
- Davis, B. L., & MacNeilage, P. F. (1994). Organization of babbling: A case study. *Language and Speech*, *37*, 341–355.
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, *38*, 1199–1211.
- Davis, B. L., & MacNeilage, P. F. (2002). The internal structure of the syllable. An ontogenetic perspective on origins. In T. T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 135–153). Amsterdam, The Netherlands: Benjamins.
- Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production. A comparison of phonetic and phonological approaches to first word production. *Phonetica*, *59*, 75–107.

- Gao, M. (2008). Mandarin tones: An articulatory phonology account. (Unpublished doctoral dissertation). Yale University, New Haven, CT.
- Giulivi, S. (2007). Vowels and consonants favored co-occurrences in language development. (Unpublished doctoral dissertation). University of Florence, Italy.
- Goldstein, L. M., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge, England: Cambridge University Press.
- Goldstein, L. M., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Berlin, Germany: Mouton de Gruyter.
- Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.
- Hayes, B. (2004). Phonological acquisition in optimality theory: The early stages. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 158–203). Cambridge, UK: Cambridge University Press.
- Iskarous, K., Goldstein, L. M., Whalen, D. H., Tiede, M. K., & Rubin, P. E. (2003). CASY: The Haskins configurable articulatory speech synthesizer. In D. Recasens, M.-J. Solé, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 185–188). Barcelona, Spain: Universitat Autònoma de Barcelona.
- Jakobson, R. (1968). *Child language, aphasia and phonological universals*. The Hague, The Netherlands: Mouton.
- Kent, R. D., Mitchell, P. R., & Sancier, M. (1991). Evidence and role of rhythmic organization in early vocal development in human infants. In J. Fagard & P. H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action* (pp. 135–139). New York, NY: Elsevier.
- Koopmans-van Beinum, F. J., Clement, C. J., & Dikkenberg-Pot, I. v. d. (2001). *AMSTIVOC (AMsterdam System for Transcription of Infant VOCalizations) applied to utterances of deaf and normally hearing infants*. Paper presented at Eurospeech 2001, Aalborg, Denmark.
- Levitt, A. G., & Wang, Q. (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, *34*, 235–249.
- Lindblom, B. E. (2008). The target hypothesis, dynamic specification and segmental independence. In B. L. Davis & K. Zajdó (Eds.), *The syllable in speech production* (pp. 327–353). New York, NY: Lawrence Erlbaum Associates.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.
- Locke, J. L. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*, 499–546.
- MacNeilage, P. F. (2008). *The origin of speech*. Oxford, England: Oxford University Press.
- MacNeilage, P. F., & Davis, B. L. (1993). Motor explanation of babbling and early speech patterns. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 341–352). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- MacNeilage, P. F., & Davis, B. L. (2000a). Deriving speech from nonspeech: A view from ontogeny. *Phonetica*, *57*, 284–296.
- MacNeilage, P. F., & Davis, B. L. (2000b). On the origin of internal structure of word forms. *Science*, *288*, 527–531.
- 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.
- Maddieson, I., & Precoda, K. (1992). Syllable structure and phonetic models. *Phonology*, *9*, 45–60.
- Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and modelling* (pp. 131–149). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Matyear, C. L. (2007). Lingual co-occurrence constraints in babbling: An acoustical study. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Conference of Phonetic Sciences* (pp. 601–604). Dudweiler, Germany: Pirrot.
- Moore, C. A., & Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors? *Journal of Speech and Hearing Research*, *39*, 1034–1047.
- Moore, C. A., Smith, A., & Ringel, R. L. (1988). Task-specific organization of activity in human jaw muscles. *Journal of Speech and Hearing Research*, *31*, 670–680.

- Nam, H., Giulivi, S., Goldstein, L. M., Levitt, A. G., & Whalen, D. H. (2011). *Computational simulation of CV combination preferences in babbling*. Manuscript submitted for publication.
- Nam, H., Goldstein, L. M., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (Eds.), *Approaches to phonological complexity* (pp. 299–328). Berlin, Germany: Mouton de Gruyter.
- Nam, H., Goldstein, L. M., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America*, *115*, 2430–2430.
- Newmeyer, F. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford, England: Oxford University Press.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Oller, D. K., & Ramsdell, H. L. (2006). A weighted reliability measure for phonetic transcription. *Journal of Speech, Language, and Hearing Research*, *29*, 1391–1411.
- Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*, 1–11.
- Ostry, D. J., & Flanagan, J. R. (1989). Human jaw movement in mastication and speech. *Archives of Oral Biology*, *34*, 685–693.
- Ostry, D. J., Vatikiotis-Bateson, E., & Gribble, P. L. (1997). An examination of the degrees of freedom of human jaw motion in speech and mastication. *Journal of Speech, Language, and Hearing Research*, *40*, 1341–1351.
- Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, *92*, 3078–3096.
- Ramsdell, H. L., Oller, D. K., & Ethington, C. A. (2007). Predicting phonetic transcription agreement: Insights from research in infant vocalizations. *Clinical Linguistics and Phonetics*, *21*, 793–831.
- 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.
- Steeve, R. W., Moore, C. A., Green, J. R., Reilly, K. J., & McMurtrey, J. R. (2008). Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *Journal of Speech, Language, and Hearing Research*, *51*, 1390–1404.
- Stockman, I. J., Woods, D. R., & Tishman, A. (1981). Listener agreement on phonetic segments in early infant vocalizations. *Journal of Psycholinguistic Research*, *10*, 593–617.
- Sussman, H. M., Duder, C., Dalston, E., & Cacciatore, A. (1999). An acoustic analysis of the development of CV coarticulation: A case study. *Journal of Speech, Language, and Hearing Research*, *42*, 1080–1096.
- Turvey, M. T. (1990). Coordination. *American Psychologist*, *45*, 938–953.
- Vilain, A., Abry, C., Badin, P., & Brosda, S. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 2, pp. 2497–2500). San Francisco: University of California, Berkeley.
- Vorperian, H. K., Kent, R. D., & Lindstrom, M. J. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America*, *117*, 338–350.
- Werker, J. F. (1989). Becoming a native listener. *American Scientist*, *77*, 54–59.
- Whalen, D. H., Giulivi, S., Nam, H., Levitt, A. G., Hallé, P. A., & Goldstein, L. M. (2011). *Biomechanically preferred consonant-vowel combinations occur in adult lexicons but not in spoken language*. Manuscript submitted for publication.
- Whalen, D. H., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, *97*, 2533–2539.
- Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, *18*, 501–516.
- Zmarich, C., & Miotti, R. (2003). The frequency of consonants and vowels and their co-occurrences in the babbling and early speech of Italian children. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1947–1950). Barcelona, Spain.