

FROM IDIOSYNCRATIC PURE FRAMES TO VARIEGATED BABBLING: EVIDENCE FROM ARTICULATORY MODELLING

VILAIN Anne, ABRY Christian, BADIN Pierre, BROSDA Stefanie
*Institut de la Communication Parlée (ICP),
CNRS UPRESA 5009,
46 Avenue Félix Viallet, 38031 Grenoble Cedex 01 - France
Fax: Int + 33 - 4 - 76.57.47.10 - Phone: Int + 33 - 4 - 76.82.41.19
e-mail: avilain@icp.inpg.fr*

ABSTRACT

The "Frames then Content" theory describes babbling as the first emergence of speech with syllable-like rhythm, under the form of "pure" and "fronted frames". "Pure frames" consist of a labial closure due to a simple mandibular oscillation, without any proper control of carried articulators. Whereas the "fronted" frames associate a fronting movement of the tongue with that jaw movement, resulting in a coronal closure. We tested this hypothesis with three articulatory models each based on X-rays from a different speaker. These simulations account for concurrent labial and coronal frames, in different babbling databases, as different idiosyncratic pure mandibular frames. The resultant forms depend on: (i) one central command, (ii) the peripheral babies' morphology, (iii) biological noise, peripheral and central. The integration of this biological noise into control signals could achieve the next step in speech development, i.e. the acquisition of segmental independence, through the control of carried articulators.

1. INTRODUCTION

The starting point of our proposal was MacNeilage's "Frames then Content" theory [10], in which he made an attempt to put together in a coherent scenario behavioral and neurological data on animal vocalisation, child babbling and adult speech. His aim was to propose a new explanation for the ontogenetic and phylogenetic evolution of speech production.

We want to put his hypothesis within the framework of articulatory modelling, in order to test the validity of his view of articulatory commands and constraints.

Our point in using different articulatory models will be that they consist of individual anthropomorphic models, which have been elaborated in such a way as to individuate the degrees of freedom of the speaker while preserving the information about his own morphology. Instead of endeavouring to normalise, we will choose to take advantage of inter-individual variation. Consequently, we will be able to simulate the behaviour of one individual "baby" at a time, with all its proper characteristics, by "freezing" its degrees of freedom except for the jaw, so as to approach the situation of a "baby" controlling this oscillation of the mandibula only.

Our final aim will be to discover an ecological path for the progressive development of the control of the baby's degrees of freedom, that is for the progressive elaboration of the phonetic sound types, with canonical babbling as the starting frame, and adult segmental complexity as the ultimate goal.

2. BABBLING PATTERNS

2.1. Canonical and variegated babbling

Speech babbling rhythm appears suddenly in infants at about seven months of age, under the form of repetitive syllable-like utterances. This form of vocalisation is called "canonical babbling": "Canonical babbling can be characterised as phonation accompanied by an alternation of closed and open phases of the mouth within spatio-temporal timing patterns appropriate for adult utterance strings. [...]" (Davis & MacNeilage, [6]).

After some years of discussion, since its definition at the end of the 70s, it became well established that canonical babbling can take either a reduplicated form or a variegated one, and this from the beginning of babbling (Smith et al., [15]). Reduplicated babbling consists of a repetition of the same consonant-vowel like sequence, such as [baba] or [dada]. Whereas variegated babbling introduces a variation of either the "consonant" [bada], either the "vowel" [babe], or both of them [bade]. For the moment, not all these variational patterns can be accounted for by such a theory as "Frames then Content".

2.2. The Frame/Content theory of speech production in ontogeny

MacNeilage's Frame/Content theory defines the rhythmic and repetitive utterances of babbling as the resulting products of mandibular cyclicities (what he calls "frames"), accompanied by phonation, but without any independent control of carried articulators (i.e. the lower lip, and the different degrees of freedom of the tongue). Active control of the degrees of freedom of these articulators develops later, producing the phonetic "content" within the frames.

The high position of the jaw produces a consonant-like closure, whereas the low position yields a full resonant nucleus, which sounds like a vowel.

Hence the variations found in variegated babbling are attributable, according to MacNeilage, merely to changes in the degree of jaw elevation. Therefore the variation in consonants will mostly consist in changes of manner of articulation, say [bava], rather than place and, for vowels, in changes of height rather than place, say [babe].

But how to take into account the fact that several studies have shown that labial and coronal closures appear quite equally in baby productions, across different subjects in different languages (for a recent review, see Brosda [5])? MacNeilage's explanation for such a distribution is that bilabial frames would be "pure frames", in that they would be the direct result of jaw elevation. Whereas coronal closures would be "fronted frames", where a

forward tongue movement takes place before the onset of the mandibular movement: the tongue being preset, the following mandibular oscillation naturally results in a coronal contact. By way of analogy, velar closures are labelled "backed frames". In this way MacNeilage explains a variety of places of contact without any need for a sequential articulatory synergy, that is without changing simultaneously the control of both the jaw and the tongue.

Notice that MacNeilage's theory, even with those tongue presettings, cannot account for variations of the consonant only [bada], or of both the consonant and the vowel [bade], along a same oscillatory sequence.

3. GENERATING FRAMES WITH DIFFERENT ARTICULATORY PLANTS

3.1. Articulatory modelling

We have used three anthropomorphic individual models or "plants" based on adult French speakers, plus a model of the growth of the vocal tract from the baby to the adult. These individual models were used to test MacNeilage's hypothesised "pure frames". We were testing: (i) if bilabial contact was the unique natural resulting product of a simple mandibular oscillation; (ii) if not, what range in the variety of contacts could be obtained from this simple mandibular (central) command, given the morphological (peripheral) differences between subjects.

3.1.1. Individual models. These articulatory models are based on the X-ray films of three French speakers, obtained from the collaboration with the staff of the Institute of Phonetics in Strasbourg, using the Schiltigheim Hospital's facilities. Maeda's model [11] was elaborated from speaker Patricia Barbier, and set within the SMIP interface (Speech Maps Interactive Plant) at ICP. The Bergame model was developed from speaker Pierre Badin [2], and the Gentiane model, from Jean-Luc Schwartz [17], both at ICP.

After drawing by hand, the profile contours were digitized, and measurements referenced to a semi-polar grid were performed. Then the degrees of freedom of the models were extracted. The statistical method, inspired from Maeda's analysis [11], consisted in applying an articulatorily-driven Overall Principal Component Analysis (PCA) to the data. (i) First the effect of Jaw Height (JH) was subtracted from the whole variance of the contours around the neutral position. (ii) The PCA analysis was then run on the body of the tongue, without taking into account the apex, so as to extract two components that correspond to Tongue Body (TB) and Tongue Dorsum (TD). (iii) Another PCA run on the apex yielded the Tongue Tip (TT) component. (iv) Finally Tongue Advancement (TA), Larynx Height (LY), Lip Height (LH), Lip Protrusion (LP), and upper Lip Vertical elevation (LV), were all kept as command parameters. As a result, 9 degrees of freedom could be obtained as control parameters for the plant.

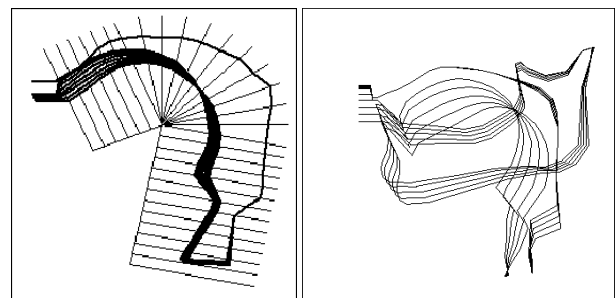
3.1.2. The Growth model. Growth comes from the Variable Linear Articulatory Model (VLAM) elaborated by Maeda (1994, unpublished), and is an articulatory model of the growing vocal tract from 0 to 21 years of age. Based on the same subject as the adult Maeda model, it has been implemented at ICP within the SMIP interface (Boë, [4]). It adds to the three models, the capacity of simulating the maturation of the dimensions of the

vocal tract with the corresponding acoustic consequences. In this study, we set the model at 6 months, approximately at the time when babbling appears. Notice that the acoustic output reflects the non-linearities in the growth of the vocal tract. But the topology of the articulator contacts remains of course unchanged. (For the sake of place we will not display here these same patterns of contacts obtained with the baby version.)

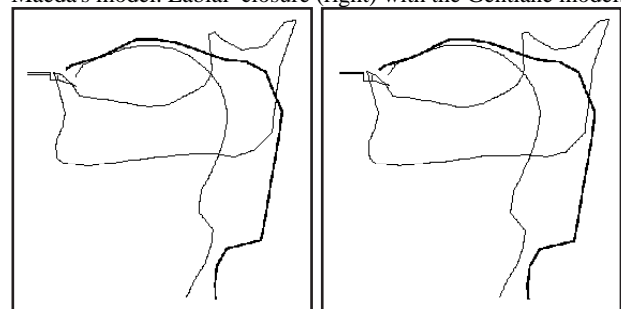
3.2. Simulation results

The simulations consisted: (i) in setting the command parameters to zero, in order to approach an ideal neutral configuration of the vocal tract; and (ii) in acting on the JawHeight parameter alone to close the vocal tract.

Figures 1-4 show the three generated frames. It appeared that Maeda's model (SMIP) clearly produced a coronal closure, say a [dada] babbling, for the adult and of course for the 6 month version as well. Gentiane yielded a labio-dental closure [vava], which could correspond to a bilabial babbling [baba], in the case where the first teeth are not yet here (this event can coincide with the babbling stage). The Bergame model produced a coronal closure, which turned closely to be to a double closure, both coronal and bilabial, when the jaw elevation was led a little further up. The explicit sequence is then [adbda]. Provided that these contacts can be quasi-simultaneous, the audible products would be either [dada]-like (if the labial release is not audible), or [bdabda]-like (if the closing phase of [d] is poorly audible).



Figures 1-2: Coronal closure (left) with the SMIP interface of Maeda's model. Labial closure (right) with the Gentiane model.



Figures 3-4: Coronal (left, the lips being not yet fully closed), or coronal-labial closure (right), with the Bergame model.

4. EVIDENCE FOR A VARIETY OF PURE FRAMES

These simulations lead us to the following conclusions.

(i) First the possible frames are coronal, labial, and coronal-labial. Thus our pool of models, in spite of being too small to constitute a representative sample, is sufficient to yield the two main consonantal place types observed in babbling databases.

(ii) Second, these three frames are obtained with a unique mandibular command.

The first consequence of results (i) and (ii) is that MacNeilage seems right at the command level. But his account of the variety of babbling types is not at all parcimonious. On the contrary our account allows to lighten the burden of a fronting movement of the tongue for coronal closures: pure frames can be coronal depending on the speaker. Like for the bilabial ones, the production of these differentiated "pure frames" does not require the intervention of any supramandibular articulator control. Such an explanation avoids MacNeilage's proposal of an unmotivated "fronting movement" of the tongue.

(iii) The third result is that the Bergame model yields the possibility of a frame with two simultaneous (or quasi-simultaneous) types of closures.

In fact, as we have suggested for this speaker, the most audible products could be either [dada] or [bdabda] sequences.

Stepping a little farther, one could make the assumption that for this speaker, we may observe a random appearance of labial and coronal products. This could account for variegated babbling of the [bdada] type, and maybe of the [bada] one.

5. BIOLOGICAL NOISE: A SOLUTION FOR CONTROLLING INDEPENDENCE OF CONTENT

For this purpose we need to introduce the notion of biological noise, both at the peripheral and central levels.

At the periphery, the inertial properties and the compliance of the carried articulators, the lower lip and the tongue, can induce asynchrony in the closure and release phases. Thus, the possibility of a double closure which is obvious in Bergame can easily be imagined to be present in the other two subjects as well. Clearly, in the SMIP, if the movement of jaw elevation is not stopped at the first closure, which is coronal, the tongue that has reached the roof of the mouth can flatten against the palate, until the lower lip reaches in turn the upper lip contact. Consequently, if the compliance of the articulators is taken into account, two closures can be performed at a time, and the release can concern either or both of them. For Gentiane, the lower and upper lip compliance (for a clear example in an 8-month-old, see Munhall and Jones [12]), can lead to a double contact pattern, yielding a variety of audible cyclicities from the [abdba] sequence.

The last step would consist in the integration of biological noise into the central control signals along the learning process. Such a possibility does not seem too far from reach, if we take seriously a recent proposal by Harris and Wolpert [8].

They introduced the neurophysiological concept of biological noise to account for a broad range of movements including eye saccade, pointing and rhythmic arm movements. With the maximum precision principle that they used to model such smooth movements, Harris and Wolpert took advantage of the knowledge that any motor command is accompanied by biological noise. That is noise in the firing of motor neurons, which is supposed to be proportional to the size of control signals. This is nicely demonstrated by the speed-accuracy trade-off captured by the well-known Fitts's law, which relates movement time directly to movement amplitude and inversely to target size. During motor skill acquisition one learns to develop strategies to carry out desired trajectories in the presence of noise by implementing this trade-off. For the baby, its motor command is rather distorted by neural noise. Only later will it learn to

integrate this neural noise into the corresponding control signals. Meanwhile the result is not always precisely the intended one.

This could explain an intra-individual variation in infant babbling, merely due to biological noise, i.e. accidental approximations in the positioning of the jaw and the carried articulators, without presupposing a precocious control of the degrees of freedom of the vocal tract.

Hence, it is not only intra-individual variation in consonant manner and vowel height, due to different amplitude of mandibular oscillation, but also changes in place of consonants, that could be accounted for by biological noise. Finally, this very notion of biological noise can account for the simultaneous appearance of reduplicated and variegated babbling at the same stage in development, though with minimal control, and it could ultimately account for the emergence of carried articulator independence from the jaw, that is in MacNeilage's words: segmental content.

6. FRAMES THEN CONTENT WITHOUT BROCA ?

The cortical networks responsible for the frames and the content remain to be elucidated. MacNeilage [10] proposed the mesial system, with the Supplementary Motor Area (SMA or PRE-SMA) as the candidate for frame generation, the lateral system, mainly Broca, being in charge of the content.

In this debate, three contradictors proposed the lateral system both for frame and content. Abry et al. [1] helped MacNeilage in discarding this lateral hypothesis by putting forward aphasiological data that had remained otherwise ignored in the controversy. At the beginning of the 80s, Poeck et al. [13] from Aachen reported 8 cases of patients suffering from global aphasia, but still producing stereotypical utterances, consisting in one recurrent pseudo-CVCV pattern, like in canonical babbling, and fully comparable to the unique "tantan" production of Broca's famous patient, Leborgne. And after the debate we discovered that Blanken et al. [3] from Freiburg-in-Brisgau had added in 1990 about 30 cases with similar recurring patterns.

As underscored by MacNeilage [10, p. 536], this set of aphasics constitutes an "insuperable problem" for those who would still hold that both frames and content depend on the lateral system, say Broca.

Poeck et al. [13] did not scan the SMA region, but Stefanuto [16] proposed that this type of aphasia could be put together with cortical stimulation data and irritative lesion data already gathered by MacNeilage to support SMA as a good candidate for frame control. While cortical stimulation data have been suspected to provoke an artifactual functioning of SMA, the irritative lesion effects (Jonas [9]) have not been questioned in the debate. In our opinion, SMA can be considered as taking part in the initiation or release of the mandibular cyclicities in speech. But the clearest point is not a positive stance in favour of SMA, instead a negative one against Broca. For Broca cannot be considered as responsible for the control of the carrier of speech, be it in babbling infants, in adults, or in aphasics regressing to babbling. For infants, this negative stance can be reinforced by neurodevelopmental data. Scheibel [14] has concluded that dendritic connections between the orofacial motor cortex and Broca's area could be not established until 12 to 15 months of age, i.e. long after the 7-month babbling emergence.

Finally it is not sure for us that Broca has to be recruited for the emergence of the independence of articulators, as proposed by MacNeilage. The recovery of the *given* degrees of freedom of the

tongue and the lips does not seem to necessitate such a specific "grammatical" device (see the forthcoming debate with Grodzinsky [7]). After all, such a device is not needed for the recovery of the anatomically *given* degrees of freedom of the hand.

In fact, Broca aphasics do not seem to have pronunciation problems with the phonology of various automatisms. Lexical automatisms, typically curses, greetings, and onomatopoeia (i.e. animal calls), all with language specific phonology. Non-lexical automatisms are essentially the pseudo-CVCV recurring utterances mentioned above. But production problems happen even for single word pronunciation, as soon as this word has flexional properties and morphosyntactic government. Thus Broca can be considered as a necessary piece in the operations of recovery and production of a grammatically flexible phonological form. Which has nothing to do with the recovery of the given degrees of freedom of our articulators.

7. CONCLUSION

The issues we addressed relate more fundamentally to the specificity of human language. Man is not the only species that uses communicative mandibular cyclicities: lipsmacks are affiliative signals in rhesus monkeys. But Man is the only one that superimposes segmental distinctions on this cyclic behaviour. So the ontogenetic issue of segmental independence questions the phylogenetic evolution of speech, according to the widespread idea that "ontogeny recapitulates phylogeny".

More specifically, what could be then the path towards recovering other degrees of freedom, i.e. towards the fulfilling of the frames with content? How did the baby develop the control of its articulators which is necessary for the next stage of speech ontogeny: segmental independence ?

Along this contribution, we became able via articulatory modelling to characterise frames as varying types of closures, according to the individual morphology of the baby, thus elucidating the real first steps of speech development. For the next challenge, phonetic "content", we now have a hypothesis, the maximum precision principle, which could be used for the learning of segmental independence through a control integrating biological noise for the jaw and the articulators it carries. The beginning of the story is that one central motor command can yield several acoustic products because of peripheral constraints and biological noise. The segmental independence can be traced back to this phenomenon, since the baby could progressively recover the control of the accidental products in this noisy relationship between the motor and the bio-acoustic levels of babbling.

ACKNOWLEDGEMENTS

In alphabetic order: Monica Baciu-Raybaudi, Gérard Bailly, Rafael Laboissière, Peter MacNeilage, Christoph Segebarth, Muriel Stefanuto, Carol Stoel-Gammon.

REFERENCES

- [1] Abry C., Boë, L.-J., Laboissière, R., Schwartz, J.-L. 1998. Emergence of a new puzzle for the theory of evolution ?, *Behavioural and Brain Sciences*, 21, 4, pp. 512-513.
- [2] Beautemps, D., Badin, P., Bailly, G., Galvan, A., Laboissière, R. 1996, Evaluation of an articulatory-acoustic model based on a reference subject, *First ESCA Tutorial and Research Workshop for Speech Production Modelling*, Autrans, France, pp. 45-48.
- [3] Blanken, G., Wallesch, C.W., Papagno, C. 1990. Dissociations of language functions in aphasics with speech automatism (recurring utterances), *Cortex*, 26, pp. 4163.
- [4] Boë, L.-J., Maeda, S. 1997. Modélisation de la croissance du conduit vocal. Espaces vocaliques des nouveaux nés et des adultes. Conséquences pour l'ontogénèse et la phylogénèse. *Journées d'Etudes Linguistiques: La voyelle dans tous ses états*, Nantes, pp. 98-105.
- [5] Brosda, S. 1998. Du babillage canonique à la naissance du contrôle des degrés de liberté des articulateurs, *Maîtrise de Sciences du Langage*, ICP, Grenoble, France.
- [6] Davis, B., MacNeilage, P. 1995. The articulatory basis of babbling, *Journal of Speech and Hearing Research*, 38, 4, pp. 1199-1211.
- [7] Grodzinsky, Y. 2000. The neurology of syntax: Language use without Broca's area, *Behavioural and Brain Sciences*, 23, 1.
- [8] Harris, C., Wolpert, D.M. 1998. Signal-dependent noise determines motor planning, *Nature*, 394, pp. 780-784.
- [9] Jonas S., 1981, "The supplementary motor area and speech emission", *Journal of Communication Disorders*, 14, pp. 349-373.
- [10] MacNeilage, P. 1998. The frame/content theory of evolution of speech production, *Behavioral and Brain Sciences*, 21, 4, pp. 499-521.
- [11] Maeda, S. 1979. An articulatory model of the tongue based on a statistical analysis, *Journal of the Acoustical Society of America*, 65, S22.
- [12] Munhall, K.G., Jones, J.A., 1998, Articulatory evidence for syllabic structure, *Behavioral and Brain Sciences*, 21, 4, pp.524-525.
- [13] Poeck K., Bleser R. de & Keyserlingk D. Graf von, 1984, Neurolinguistic status and localization of lesion in aphasic patients with exclusively consonant-vowel recurring utterances, *Brain*, 107, pp. 199-217.
- [14] Scheibel, A. 1993. Dendritic structure and language development, in B. Boysson-Bardies et al. (eds), *Developmental Neurocognition: Speech and face processing in the first year of life*, 51-62, Kluwer Academic Publishers, Netherlands.
- [15] Smith, B.L., Brown-Sweeney, S., Stoel-Gammon, C. 1989. A quantitative analysis of reduplicated and variegated babbling, *A First Language*, 17, pp.147-153.
- [16] Stefanuto, M. 1997. Apport de la neurocognition comparée, développementale et pathologique dans l'étude des tendances universelles phonologiques, DEA Sciences du Langage, ICP, Grenoble, France.
- [17] Vilain, A. 1997. Un nouveau modèle articulaire pour la synthèse et le contrôle robotique de la parole: Gentiane, DEA Sciences du Langage, ICP, Grenoble, France.