Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, Spain. August 3–9, 2003. Universitat Autònoma de Barcelona.

1297

Emergence of discrete gestures

Louis Goldstein
Department of Linguistics, Yale University
Haskins Laboratories, New Haven, CT
E-mail: louis.goldstein@yale.edu

ABSTRACT

In order for constriction actions (gestures) produced by vocal tract articulators to function as phonological primitives, they must be discrete. Two potential sources of discreteness are proposed here: the distinct organs of the oro-facial anatomy that can produce vocal tract constrictions and the differentiation of a given constricting organ's behavior into distinct actions through mutual attunement among the members of a community. Organ distinctions appear to be more fundamental in that they are already respected at birth, while differentiation of organ action requires experience and may differ from language to language, as is shown in a new simulation presented here. These differences are predicted to affect the course of phonological development: systematic, adult-like deployment of organs should precede differentiation of organ action in children's early words. Results of a new study support this prediction.

1. INTRODUCTION

The fundamental insight of phonology is that speech can be decomposed into a small number of atomic units that can recombine in different arrangements to form the words of a language. The failure to find these discrete units in measurements of the speech process led some to the view that the phonological and physical descriptions of speech are incommensurate [1]. Articulatory phonology [2,3,4] has attempted to reconcile these descriptions by hypothesizing that while the *products* of speech production (articulator movements, airflow, acoustics) are continuous and context-dependent, the *act* of speech production itself can be decomposed into discrete, dynamical units of action, or *gestures*. Because they are discrete, gestures can function as units of contrast and combination. Because of their dynamical nature, they can be linked in a principled way to the continuous physical structure of speech.

If speech is indeed composed of discrete action units (and this is by no means uncontroversial), it is important to ask how such units arise (in the child and/or in the evolution of language). What is the basis for their discreteness?

2. VOCAL ORGANS

The vocal tract can be viewed as harboring a number of independent constricting devices, or organs: lips, tongue tip, tongue body, tongue root, velum, and larynx. They are independent (even though mechanically coupled) in the sense that a constriction can be formed by one of them, without necessarily causing a constriction in any of the others[5]. A gesture is hypothesized to be a constriction action of one of these organs [6], and thus, gestures of distinct organs embody a discrete difference. A lip gesture vs. a tongue tip gesture is a discrete difference because different body parts act to constrict the vocal tract.

While the ability of the constricting organs to act independently is a necessary condition for them to function as combinatoric atoms, it is not sufficient to explain why body part differences *count* as discrete, potentially informational differences. What leads us to view our body parts as distinct objects?

The partitioning of the oro-facial system into functionally and anatomically distinct organs is apparently a very basic part of human biology. Evidence for this can be found in the research on facial mimicry in infancy. Meltzoff & Moore [7,8] have shown that even human neonates are capable of facial mimicry—remarkably so, as infants cannot see their own faces, nor feel proprioception from the model's face. What it means for infants to mimic facial displays (as described in [8]), is that they move the *same organ* as the model (e.g., lips, tongue) and position it in the same relation to the rest of the body. In fact, they may not succeed in positioning the organ correctly on their first attempt, but they appear to always move the right organ.

Meltzoff & Moore propose a model of facial mimicry in which an infant is able to innately identify her own organs (that she feels) with the organs of the model (that she sees), within a common representational framework. If this innate ability extends to the speech organs (which overlap, in part, with the facial organs investigated in the mimicry experiments) then this provides a firm basis for the fact that actions of distinct speech organs count as discrete differences. In addition, the ability to identify one's own organs with those of the "other" is exactly what would be required to use discrete organs as the basis for a communication system. Based on these considerations, Studdert-Kennedy has argued that the "particulation" of the vocal tract could be the evolutionary source of discreteness in language [9, 10].

Some cautions do need to be considered in extending the Melzoff & Moore organ model to the speech organs. First,

ISBN 1-876346-48-5

most of the information about the behavior of speech organs (during speech) comes through the auditory system rather than the visual system (*all* of the information about some speech organs, e.g., the velum and the glottis, must come that way). Thus, if the extension holds, it would predict that infants could mimic the motions of organs for which they receive auditory rather than visual information. For example, an infant should produce some movement of the lips if she hears a model perform a nonspeech action of the lips that produces a sound, e.g., a raspberry. The fact that infants can integrate audio and visual information about speech [11], suggests that they may well be able to do this.

Second, some of the distinct speech organs may not be differentiated in early infancy, for example, the tongue tip vs. tongue body. It would be difficult to evaluate the claim that they are distinct organs at that time, as it is not clear that a very young infant could move the tongue tip independently of the tongue body, even if she "knew" it was as distinct organ. So there may be differences among organs in their developmental course.

Of course, not all contrasting gestures employ distinct organs. Words like "tick," "sick," and "thick" all begin with gestures of the tongue tip organ. Nonetheless there is some evidence that between-organ contrasts are primary within the phonologies of languages. Between-organ consonant contrasts employing lips, tongue tip, tongue body, and velum organs are close to universal [12], while several within-organ contrasts (e.g., /t/-/θ/) are quite rare. Evidence from phonological alternations also supports this view. Features corresponding to organ contrasts (Place, Nasal, Larynx) are near the top of feature geometry hierarchies, designed to capture the behavior of features in alternations [13].

While between-organ contrasts may be primary, within-organ contrasts do exist, and they can by described as contrast in the values of the constriction goals, for example the location and degree of the constriction. Going back to the model of Meltzoff & Moore, these are differences in organ *relation*. However, parameters like degree and location of constriction are continua. How are these continua partitioned into discrete regions that can serve as goals for contrasting gestures?

3. WITHIN-ORGAN ATTUNEMENT

When gestures are used as part of a communication system, there must be agreement among the members of the speech community as to what the gestures are. Since every member has the same organs and an innate ability to recognize the identity of those organs in the self and others, this agreement is guaranteed in the case of the organ employed for a given gesture. As regards the values of the gesture's constriction goals, agreement must come from elsewhere. The members of the community must attune their constriction actions to one another's. In previous work it has been shown [4], that under certain conditions, this process of attunement can also automatically partition a constriction continuum into discrete, potentially contrasting intervals.

The attunement process has been modeled as a "game" in which two computational agents interact ([14], cf. related work in [15]). At the onset of the game, each agent has an equal probability of producing all of the values along a constriction continuum (e.g., constriction degree, *CD*). On each trial of the game, both agents produce some value of *CD* at random, weighted by the current probabilities associated with each value (e.g., $P(CD_i)$). Each agent then recovers the partner's production, and compares the produced value to the recovered one. If they match, then $P(CD_i)$ is incremented.

Under certain recovery conditions, it has been shown that the agents will partition the continuum into distinct regions or modes. These conditions include (1,2) below:

(1) Recovery is assumed to be noisy, so a range of values on either side of the produced value is recovered.

(2) The mapping from constriction value to acoustic parameter is assumed to be a nonlinear step function, like that seen in Fig.1 (after Stevems[16]).

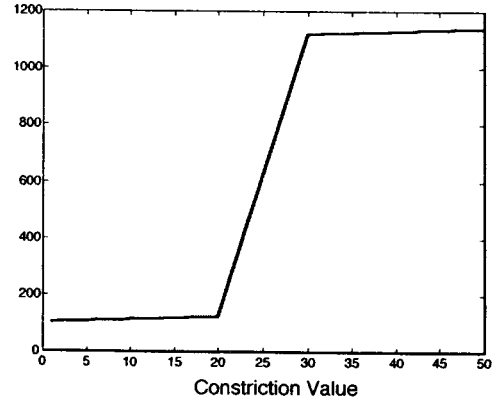

Figure 1. Nonlinear relation of constriction value to acoustic output (ordinate) used in agent simulations.

When these conditions are satisfied, attunement partitions the *CD* values into intervals corresponding to the stable regions in Fig.1, e.g., stops and fricatives.

However, we know that not all constriction continua are associated with a step-like mapping like that in Fig. 1. Constriction location (CL) of tongue tip gestures (e.g., dental vs. alveolar stops in Australian languages) or of tongue body gestures (velar vs. uvular stops) could be such examples.

To see compare how partitioning a continuum into contrasting modes could be achieved both with and without a step function mapping, the following attunement simulation was undertaken. Each of two computational agents produced a pair of different gestures (A and B). For

ISBN 1-876346-48-5 © 2003 UAB

each gesture, the value of the constriction parameter was chosen at random from a sequence of 50 intervals, weighted by the probability associated with producing that interval (probabilities were all equal at the outset). Each agent then compared the value it produced for gesture A with the recovered version of the partner's production of gesture A, and if they matched, two things occurred: the probability of producing that constriction value for A was incremented (rewarded), and the probability of producing that same value for gesture B was decreased (punished). This coupling of A and B was designed to simulate the idea that the agents are attempting to produce A and B "differently" (contrastively). In fact, not only was the matched value of A punished for gesture B, but also any nearby values that could be recovered as that value. Two different maps relating constriction value to (simulated) acoustic value were employed: the nonlinear step function in Fig. 1 and a linear function with slope 1 (which is also the slope in the "stable" regions of Fig. 1). Acoustic noise was set at 4 units, so a given produced constriction value would match any of the partner's productions whose acoustic output was within 4 acoustic units of the produced value. The simulation was repeated 100 times, representing the self-organization of 100 different "languages."

Results showed that for both linear and nonlinear maps, the agents converged in producing a pair of contrasting values of the constriction parameter. The difference between the maps was in the uniformity of the converged state across the 100 "languages." For the linear map, gesture pairs were evenly distributed over the 50 posssible constriction values. There was no preference for particular pairings (apart from the fact the nature of the "punishment" meant the pair of gestures had to be at least 8 units apart). For the nonlinear map, 77% of the languages contrasted values from the two different stable regions of the map. These results appear to correspond well to the cross-linguistic differences in CD and CL. While most languages have contrasts between stops and fricatives, or fricatives and glides, the exact nature of CL contrasts appears to be quite variable [17], as is the value of these parameters in languages in which they do not contrast [18]. Attunement under different articulatory-acoustic conditions can account for these differences.

4. PHONOLOGICAL DEVELOPMENT

The hypothesis that the source of discreteness is different for between- vs. within-organ contrasts makes predictions about the course of phonological development. Since neonates can already identify oro-facial organs, we would expect between-organ contrasts to emerge quite early in development. Within-organ contrasts require attunement to the language environment and thus should take longer to emerge. Reanalyses of some existing data from individual babies has provided some support for this prediction [19].

To test this hypothesis more systematically, recordings of six American English-learning infants collected by Bernstein-Ratner [20] and available in the CHILDES database [21] below were analyzed[1]. The recordings of child-parent interactions took place in an environment with the same set of play objects, so that the target words for the infant productions could often be inferred fairly unambiguously from context and were the same from infant to infant. Ages at the time of the recording ranged from 1:1 to 1:9.

Infants' words for which adult targets could be inferred were extracted, randomized, and presented to a panel of ten listeners who transcribed the English consonant that they heard at the beginning of each word. A total of 174 words were transcribed.

Transcriptions were compared to the adult targets, and errors were analyzed. The percentage of errors sharing various gestural properties was calculated, for example the number of errors sharing the correct oral constricting organ (lips, tongue tip, tongue body). To provide a chance baseline for significance testing, the transcriptions and the adult forms were re-paired randomly 100 times. The number of times out of these 100 that the sharing percentage was as high as that in the correctly paired data is taken as significance measure.

Table 1 shows the results pooled over the six infants. Column 1 gives the percentage of errors that shared the indicated gestural property; column 2, the average percentage of errors that shared the property in the randomized data; column 3, the number of randomizations yielding a percentage as high as the data, and column 4, the number of infant words analyzed

	1	2	3	4
Oral Constrictor	49.93	32.91	0	161
Velum	80.34	78.49	7	174
Glottis	63.45	48.55	0	174
Constriction Degree	55.25	60.66	99	174

Table 1. Results of analysis of errors on initial consonants in infants' words. (See text).

Data indicate that infants employ the correct oral constricting organ at the beginning of words. Separate analyses of the individual infants were all significant. For the velum and glottis organs, infants tended to match their states also, but not as completely as the oral constriction organs. The overall results for the velum were not quite significant, and individual infants varied as to whether they showed significance or not. In contrast to the results for the organs, the infants did not show any tendency to match the constriction degree, did not show any tendency to match the constriction degree of the adult form (coded as stop, fricative, glide). One of the six infants did show a significant tendency to share CD, but she contributed a total of only 8 usable utterances.

[1] Work in collaboration with Jacob Taylor

ISBN 1-876346-48-5 © 2003 UAB

These results support the organ hypothesis. From infants' the earliest words, they are using distinct organs in much the same way adults do. However, the process of attunement has not yet proceeded to the point where constriction degree is distinguished systematically.

It is possible to see organ-related developmental differences in speech perception as well. The classic experiments (e.g.,[22]) showing decline in discriminability for non-native contrasts at around 10 months of age employed within-organ contrasts (dental vs. retroflex coronals; velar vs. uvular dorsals). Recent work [23] suggests that the ability to discriminate between-organ contrasts may not decline in the same way. Indeed, we may retain into adulthood our fundamental ability to perceive as distinct the actions of the distinct body organs.

ACKNOWLEGMENTS

This work was supported by NICHD Grant HD-01994 and NIDCD Grant DC-00408 to Haskins Laboratories.

REFERENCES

[1] C. Hockett, *A Manual of Phonology.* Bloomington, Indiana: Indiana University Press, 1955.

[2] C. Browman and L. Goldstein, "Articulatory Phonology: An Overview", *Phonetica,* vol. 49, pp. 155-180.

[3] C. Browman and L. Goldstein, "Dynamics and articulatory phonology, in *Mind as Motion: Explorations in the Dynamics of Cognition,* R. Port and T. van Gelder, Eds., pp. 175-193. Cambridge MA: MIT Press, 1995.

[4] L. Goldstein and C. Fowler. "Articulatory Phonology: a phonology for public language use." In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* A. Meyer and N. Schiller, Eds. Berlin: Mouton, in press.

[5] M. Halle, "On distinctive features and their articulatory implementation," *Natural Language and Linguistic Theory,* vol. 1, pp. 91-105, 1983.

[6] E. Saltzman and K. G Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological Psychology,* vol. 1, pp. 333-382, 1989.

[7] A.N. Meltzoff and M.K Moore, "Imitation of facial and manual gestures by human infants," *Science,* vol. 198, pp. 75-78, 1977.

[8] A.N. Meltzoff and M.K Moore, "Explaining facial imitation: a theoretical model," *Early Development and Parenting,* vol. 6, 179-192, 1997.

[9] M. Studdert-Kennedy, "The particulate origins of language generativity," in *Approaches to the Evolution of Language,* J. Hurford, M. Studdert-Kennedy,. and C. Knight, Eds. pp. 202-221. Cambridge: Cambridge University Press, 1998.

[10] M. Studdert-Kennedy and L. Goldstein, "Launching Language: The Gestural Origin of Discrete Infinity," in *Language evolution: The States of the Art,* M. Christiansen and S. Kirby, Eds. Oxford: Oxford University Press, in press.

[11] P. Kuhl, and A.N. Meltzoff, "The bimodal perception of speech in infancy," *Science,* vol. 218, pp. 1138-1141, 1982.

[12] I. Maddieson, *Patterns of Sounds,* Cambridge: Cambridge University Press, 1984.

[13] J.J. McCarthy, "Feature geometry and dependency: A review," *Phonetica,* vol. 45, pp. 84-108, 1988.

[14] C. Browman and L. Goldstein. "Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures," *Bulletin de la Communication Parlée,* vol. 5, pp. 25-34, 2000.

[15] B. DeBoer, "Self-organization in vowel systems," *Journal of Phonetics,* vol. 28, pp. 441-465, 2000.

[16] K. N. Stevens, "On the quantal nature of speech." *Journal of Phonetics,* vol. 17, pp. 3-45, 1989.

[17] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages,* Oxford: Blackwells, 1996.

[18] S. Dart, "Comparing French and English coronal consonants." *Journal of Phonetics,* vol. 26, pp. 71-94, 1998.

[19] M. Studdert-Kennedy, "Mirror neurons, vocal imitation, and the evolution of particulate speech," in *Mirror Neurons and the Evolution of the Brain and Language,* M. Stamenov and V. Gallese, Eds., . pp. 207-227. Amsterdam: John Benjamins, 2002.

[20] N. Bernstein-Ratner, "Phonological rule usage in mother-child speech." *Journal of Phonetics,* vol. 12, pp. 245-254.

[21] B. MacWhinney, *The CHILDES project: Tools for analyzing talk. Third Edition,* Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[22] J.F. Werker and R.C. Tees, "Cross-language speech perception: evidence perceptual reorganization during the first year of life." Infant Behavior and Development, vol. 7, pp. 49-63, 1984.

[23] C.T. Best and G. W. McRoberts, "Infant perception of nonnative contrasts that adults assimilate in different ways. *Language and Speech.* in press.

ISBN 1-876346-48-5