Experimental Evidence for an Effect of Vocal Experience on Infant Speech Perception

Marilyn M. Vihman and Satsuki Nakai
University of Wales, Bangor
E-mail: m.vihman@bangor.ac.uk, s.nakai@bangor.ac.uk

ABSTRACT

We tested for an articulatory effect on perception by bimonthly recordings of monolingual English and Welsh children, from age 10.5 to 12 months. The infants were tested at 12.5 months on closely matching lists of nonword stimuli constructed to highlight one of two consonants of comparable input incidence (English /t/ vs. /s/; Welsh /b/ vs. /g/). Listening times were in inverse correlation with the children's relative frequency of use of the pair of consonants ($p<.05$), a novelty effect. The findings demonstrate an effect of motoric practice on infant speech perception. We speculate that production may support the establishment of more robust lexical and phonological representations.

1. INTRODUCTION

Locke [1] was perhaps the first to suggest that the direction of influence in the formation of phonological representations for early words might be from production to perception, or the reverse of what is usually assumed:

> If infants are predisposed to project innate movement patterns onto standard words ... only *contradictory* information may be needed from perceptual analyses for the child to avoid unintelligible results. A fairly crude analysis of his own sounds in relation to environmental sounds would be sufficient, and there would be, therefore, *a motor basis to speech perception - and internal representations - from the very start ...* (p. 245; our emphasis).

Given what is now understood about the role of implicit distributional learning in infants' growth of knowledge about their language in the second half of the first year [2], Locke's words take on a somewhat different meaning. Despite ample demonstration, over the past ten years or more, that infants display considerable sensitivity to and implicit knowledge of the distributional patterning of their language within the first year of life [3], it is likely that such learning supplies only a foundational level of phonological knowledge. When children advance from passive intake to production alongside perception of speech they lay down phonological representations at a new level. A similar developmental reorganization is suggested by Stager and Werker's recent experimental findings [4], which show that the task of relating form and meaning requires a new kind of processing with a consequent regression in phonological discrimination. However, children with larger production vocabularies proved better able to make the fine phonemic discrimination tested in a word learning task at 14 months [5]. We discuss below the role of production on memory for words and, by extension, on more general phonological knowledge or skills.

The primary goal of the study reported here was to test the idea that infants' emerging capacity to produce speech-like vocal sequences - 'canonical' CV-sequences, which reliably emerge between about 6 and 8 months of age [6] - provides them with a basis for mapping adult speech onto an internal representation (of their own babbling) that is motoric and proprioceptive as well as auditory, since the child's vocal production is necessarily heard as well as 'felt'. The 'articulatory filter' hypothesis [7, 8] is based on the idea that this mapping will serve as a bridge from the more fleeting (implicit) distributional representations of input speech to more stable representations for word forms that match child production patterns, providing a subset of more deeply encoded forms that can be retrieved at will and produced in appropriate contexts (the basis for the first production vocabulary). It is thus possible to suggest that the origins of voluntarily accessible representations may be found in vocal production, while speech perception alone (without production) provides a rich data-base for distributional learning, including knowledge about language-specific word units, prosodic cues to segmentation, and so on [9], and may thus be necessary but not sufficient as a basis for word production.

2. METHOD

A longitudinal production study was designed to test whether the relative frequency, in each infant's production, of two consonants of roughly the same input frequency in the ambient language would affect his or her relative orientation time towards nonwords consisting of these same consonants in a perception test. By establishing each infant's production patterns and then using a single contrasting pair of consonants in the perception test for each language we sought to examine directly the effect on perception or attention of familiarity based on production. Accordingly, production data were collected fortnightly from 10.5 - 12 months with both English and Welsh infants, followed by a perception test at 12.5 months.

To determine consonantal incidence for Welsh in both infant vocalizations and input speech we analyzed half-hour transcriptions of five mother-child sessions recorded

ISBN 1-876346-48-5 © 2003 UAB

in an earlier study when the infants were 12 months old [10]. For consonantal incidence in English infant vocalizations an earlier study of nine infants acquiring American English was used [11], while English input frequencies were based on [12]. At least one of each pair of consonants had been found to be frequent in production in these earlier analyses of infant vocalisations while within each language the pair of consonants chosen were of comparable input frequency.

For English, /t/ was paired with /s/; for Welsh /b/ was paired with /g/. Six disyllabic nonwords of the form CVCV were constructed for each language with each of the two target consonants as the only fully consonantal supraglottal. The second C slot was filled either by a glottal (/h, ˀ/), a glide (/w, j/), or a repeat of the target consonant (Table 1). The English stimuli were recorded by a female speaker of Southern British English, the Welsh stimuli by a female Welsh speaker from North Wales. Acoustic analyses established that there were no significant prosodic differences between the contrasting lists in each language.

English

spelling	IPA	spelling	IPA
tetty	/tɛti/	sessy	/sɛsi /
tooyer	/tujə/	sooyer	/sujə/
tawey	/tawi/	sawey	/sawi/
tutter	/tʌtə/	susser	/sʌsə/
teehy	/tihi/	seehy	/sihi/
tetter	/tɛtə/	sasser	/sæsə/

Welsh

spelling	IPA	spelling	IPA
bewa	/bewa/	gewa	/gewa/
bawy	/bawɨ/	gowy	/gowɨ/
boia	/boja/	gwia	/guja/
bwba	/buba/	giga	/giga/
bibu	/bibɨ/	gagu	/gagɨ/
bybu	/bəbɨ/	gygu	/gəgɨ/

Table 1: Nonword stimuli.

For production, at each data collection point a half-hour unstructured interaction between infant and mother was recorded using a DAT recorder; the infant wore a wireless microphone hidden in a specially designed vest. Infant vocalisations were transcribed by two transcribers. We included in the subsequent analyses only those consonants agreed upon by both transcribers in terms of place and manner of articulation (i.e., excluding differences as to voicing, which is neither well controlled at this age nor reliably transcribed in general: [13]).

For the perception test we followed the procedure detailed in [14], with a few changes. The number of test trials was reduced from six to four (one list from each side x 2 lists), since the stimuli in each list were very similar and rapid habituation to the stimuli was therefore expected. Each of the six words were repeated once in every trial (unless the infant had terminated the trial by looking away from the speaker for more than 2 seconds); thus, 12 words were presented during each trial. Finally, as a further precaution against infant habituation, a different pair of nonwords was used during training.

3. RESULTS

All 27 English infants (15 males) and 26 Welsh infants (13 males) completed the study. Three additional Welsh infants were tested but excluded when their parents were found to use English with them during the recording sessions (monolingual infants had been recruited based on parental report of language use in the home).

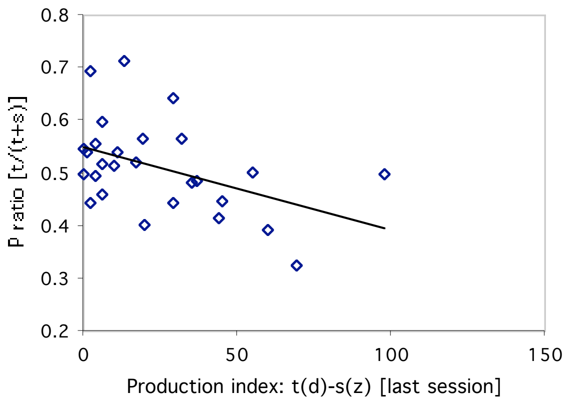
In each infant's vocalisations at each data collection point we tallied the incidence of each of the two consonants contrasted in the perception test, disregarding voicing (i.e., the incidence of [t] (and [d]) vs. [s] (and [z]) and of [p] (and [b]) vs. [k] (and [g]). We tallied each consonant separately (1) as produced in any phonotactic position (whether onset, coda, or as a syllabic nucleus) or (2) in syllable onset position only, and (3) for all four sessions or (4) for the last session only. There were thus four kinds of summary figures for each consonant, reflecting production

(a) in all positions in any session,
(b) in onset position in any session,
(c) in all positions in the last session,
(d) in onset position in the last session.

To test the correlation between production and perception data two types of indices were developed for both production and perception. For each of the four sets of *production* summary figures, the figures for the contrasting consonants [e.g. (t and d) vs. (s and z)] were combined to produce an index of the degree of the infant's production preference for one consonant (t and d) over the other (s and z) in two ways: (t+d)-(s+z) (Production difference) and t+d/(t+d+s+z) (Production proportion). Pearson product correlation tests were performed on these production preference indices in relation to the two *perception* preference indices: the Perception ratio (orientation time towards one list [list A] divided by the sum of orientation times towards the two contrasting lists [lists A and B], or total orientation time) and the Perception difference (orientation time for list A minus orientation time for list B).

Our results show that, for both languages, the Pearson correlation coefficients indicating the relationship between perception and production indices were negative, regardless of the indices used. That is, the greater the infant's production preference for one consonant over the other, the shorter his or her relative orientation time to the list containing this consonant. When both languages are considered, the best correlation was found to obtain between the Production difference between the number of consonants produced in all positions during the last session ((c) above) and the Perception ratio (Fig. 1). When the two language groups' correlation coefficients for the two indices are combined using Sachs' formula [16], we obtain: r = -.38, n = 43, p < .05.

English



Welsh

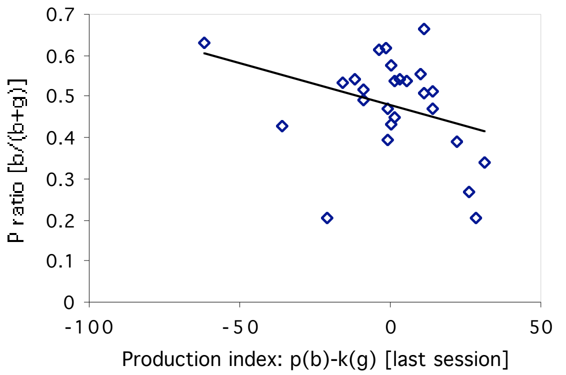


Figure 1: Production - perception indices. (Production differences vs. Perception ratios)

Note that the incidence of the consonant contrasts tested differ in the two languages. For English, one consonant type [t] (and [d]) is highly common while the other ([s] and [z]) is rare in infant vocalizations. This is reflected in the production index shown in Figure 1, in that the difference score is entirely positive ([s] and [z] are never produced more often than [t] and [d]). In Welsh, on the other hand, both [p] (and [b]) and [k] (and [g]) are produced frequently by some infants, resulting in a difference score that may be either positive or negative.

4. DISCUSSION

The fact that the perception results correlated more strongly with the production *difference* than with the production *ratio* suggests that the absolute frequency of consonant production may have a bearing on perception performance (see also [17]). On the other hand, the fact that perception and production are more strongly correlated on the basis of the perception *ratio* than the perception *difference* suggests that the *absolute* orientation time for each consonant in the perception test may in part reflect individual infants' general attention span.

Finally, the fact that the production and perception preferences are *negatively* correlated could be interpreted in at least two ways (regardless of whether infant familiarity with a particular consonant is taken to be purely perceptual, based on an increase in auditory experience due to the addition of the infant's own production to overall input frequency, or is taken to be motoric as well as perceptual, following [1] and the hypothesized articulatory filter [7, 8]). The negative correlation could be an indication of faster processing of familiar sounds. This is the interpretation given, for example, to the finding that 4-month-old monolingual children exposed to either Spanish or Catalan orient longer to stimuli from the unfamiliar than from the familiar language when Spanish and Catalan are contrasted [17].

Alternatively, the time-course of familiarization proposed by Hunter and Ames [18] may provide a framework for understanding infants' longer orientation to the consonant that they did not (or did not often) produce in the last recording session preceding the test session. Hunter and Ames propose that infants progress in a systematic fashion from a familiarity to a novelty preference for perceptual stimuli of any kind, moving increasingly quickly through the proposed sequence with age, although the nature of the particular stimuli (auditory, visual, motor, etc.) and infants' experience with it will affect the progression as well.

Based on this model, we might assume that by 12 months the infants have attained sufficient familiarity with their own well practised consonants to prefer to sample the novel stimuli in greater depth, giving more attention to relatively unknown or unpractised consonants. Based on this interpretation, sampling production at an *earlier age* might yield longer orientation times to the practised consonants, given their relatively short time in the infants' repertoire. Obtaining sufficient production data might be difficult, however, since even at 10.5 months infants tend to vocalise a good deal less, as a group, than at 12 months. On the other hand, infants tested at 12.5 months on contrasting lists of *known words* embodying the more vs. less practised consonants might orient longer to the more practised consonants as a result of the greater cognitive load of lexical processing [5, 19]. A serious methodological difficulty would be posed here, however, by the paucity of known words at this age and the consequent difficulty of controlling for phonotactic complexity.

5. CONCLUSIONS

This study produced results that are compatible with the initial hypothesis of a production effect on perception. To our knowledge, they constitute the first evidence that infants are, to some degree, affected in their listening to speech by the vocal patterns that they themselves are producing. Evidence from early word production has long

ISBN 1-876346-48-5 © 2003 UAB

been available to show that first words tend to be relatively accurate [20] and 'selected' to fit with the infant's preexisting babbling patterns [11, 21], suggesting that the infant's own vocal patterns focus attention on particular words and phrases in the input. Our findings support the idea that memory for selected words may be enhanced by emergent infant production capacities. This could be a critical factor in converting implicit distributional knowledge of the ambient language into the 'item learning' of particular words and phrases which begins by early in the second year of life [22].

Furthermore, we speculate that word learning based on production as well as perception stabilizes phonological representations, thereby creating more memorable anchors for referential word learning. Such an effect has now been demonstrated in both one-year-olds, just beginning to form an initial referential vocabulary [16], and three- to six-year-olds, whose learning of novel word-referent pairings is affected by 'phonological probability' (which is arguably related to their existing productive phonotactic repertoire, though this was not tested) [23]. Both [16] and [23] understand this link in terms of a limited capacity account of child resources, such that novel words composed of familiar sound sequences facilitate phonological processing, which in turn "may allow greater allocation of resources to lexical and semantic processing…" [23, p. 1333]. We suggest that productive knowledge of sound patterns provides a more solid or stable basis for word learning – and thus for advances in phonological ability - than does implicit distributional knowledge based on perception alone.

ACKNOWLEDGMENT

This study was funded by the UK Economic & Social Research Council (R000238236 and R000237087). We thank Pamela Martin for collecting and transcribing the data as well as running the experiments.

REFERENCES

[1] J. L. Locke, "Speech perception and the emergent lexicon", *Language Acquisition* (2d ed.), P. Fletcher and M. Garman, Eds., pp. 240-250. Cambridge: The University Press, 1986.

[2] J. R Saffran, R. N. Aslin and E. L. Newport, "Statistical learning by 8-month-old infants", *Science*, vol. 274, pp. 1926-1928, 1996.

[3] P. W. Jusczyk, *The Discovery of Spoken Language,* Cambridge MA: MIT Press, 1997.

[4] C. L. Stager and J. F. Werker. "Infants listen for more phonetic detail in speech perception than in word-learning tasks," *Nature*, vol. 388, pp. 381-382, 1997.

[5] J. F. Werker, C. T. Fennell, K. M. Corcoran and C. L. Stager. "Infants' ability to learn phonetically similar words: Effects of age and vocabulary size," *Infancy*, vol. 3, pp. 1-30.

[6] D. K. Oller, *The Emergence of the Speech Capacity,* Mahwah NJ: Lawrence Erlbaum, 2000.

[7] M. M. Vihman, "Ontogeny of phonetic gestures," *Modularity and the Motor Theory of Speech Perception*, I.G. Mattingly and M. Studdert-Kennedy, Eds., pp. 69-84. Hillsdale NJ: Lawrence Erlbaum, 1991.

[8] M. M. Vihman, "Variable paths to early word production," *Journal of Phonetics,* vol. 21, pp. 61-82, 1993.

[9] J. L. Morgan and K. Demuth, Eds., *Signal to Syntax.* Mahwah NJ: Lawrence Erlbaum, 1996.

[10] M. M. Vihman, "Prosodic development in Welsh." End of award report, ESRC Award R000237087, 2000.

[11] M. M. Vihman, M. A. Macken, R. Miller, H. Simmons and J. Miller, "From babbling to speech," *Language*, vol. 61, pp. 395-443, 1985.

[12] M. Mines, B. Hanson and J. Shoup, "Frequency of occurrence of phonemes in conversational English," *Language and Speech*, vol. 21, pp. 221-241, 1978.

[13] M. A. Macken, "Aspects of the acquisition of stop systems", *Child Phonology, I: Production* G. Yeni-komshian, J. F. Kavanagh and C. A. Ferguson, Eds., pp. 143-168. New York: Academic Press, 1980.

[14] D. Kemler Nelson, P. W. Jusczyk, D. Mandel, J. Myers, A. Turk and L. Gerken, "The headturn preference procedure for testing auditory perception," *Infant Behavior and Development*, vol. 18, pp. 111-116, 1995.

[15] L. Sachs, *Applied Statistics: A Handbook for Techniques,* Heidelberg: Springer-Verlag, 1982.

[16] L. McCune and M. M. Vihman, "Early phonetic and lexical development," *JSLHR*, vol. 44, pp. 670-684, 2001.

[17] L. Bosch and N. Sebastian-Galles, "Native language recognition abilities in 4-month-old infants from monolingual and bilingual environments," *Cognition*, vol. 65, pp. 33-69, 1997.

[18] M. A. Hunter and E. W. Ames, "A multifactor model of infant preferences for novel and familiar stimuli," *Advances in Infancy Research*, C. Rovee-Collier and L. P. Lipsitt, Eds., vol. 5, pp. 69-95, 1988.

[19] J. F. Werker and C. L. Stager, *Papers in Laboratory Phonology V,* M. B. Broe and J. B. Pierrehumbert, Eds., pp. 181-193. Cambridge: Cambridge University Press, 2000.

[20] C. A. Ferguson and C. B. Farwell, "Words and sounds in early language acquisition," *Language*, vol. 51, pp. 419-439, 1975.

[21] C. A. Ferguson, D. B. Peizer and T. A. Weeks, "Model-and-replica phonological grammar of a child's first words," *Lingua*, vol. 31, pp. 35-65, 1973.

[22] M. M. Vihman, S. Nakai and R. A. DePaolis, "Getting the rhythm right", *Papers from the 8th Conference on Laboratory Phonology,* L. Goldstein, K. Best and S. Anderson, in press.

[23] H. L. Storkel, "Learning new words," *JSLHR*, vol. 44, pp. 1321-1337, 2001.