PERCEPTION AND PRODUCTION IN EARLY L1 ACQUISITION AND SOME THEORETICAL IMPLICATIONS

Henning Wode
English Department, University of Kiel

ABSTRACT

The relationship between perception and production is complex, and it extends well beyond the phonological component of lexical items and how to learn them. The impact of perception needs to be explored also with respect to, for example, phylogeny, language change, the imperfections in early L1 production, or various issues of phonetic/phonological theory, such as the biological basis for phonological units, language specificity, or the impact of perception on the structure and the functioning of sound systems whatever their state of development. I rely primarily on perceptual data on human neonates, infants before the onset of speech, and non-human animals as well as on new production data for L1 German to speculate about some of these issues. One major conclusion will be that distinctive features, phonemes, and their typology are rooted in perception rather than production.

1. INTRODUCTION

Both perception and production are required for sound systems to be learnable and to function the way they do. Most current phonological theories, however, are still heavily biased towards production, and most mainstream acquisition researchers have only recently begun to take perception into consideration. The recent advances in speech perception research, in particular, with respect to neonates, pre-speaking infants, and non-human animals, make it clear that the impact of perception has been grossly underestimated. In fact, it now appears that perception rather than production is primary in the ontogeny, phylogeny, and functioning of sound systems. For example, consider questions like the following: Why is only a limited number of sound contrasts used for phonemic purposes in the languages of the world? Why only those that do get used, although the human vocal tract allows for the production of many other sounds? Given the familiar segmentation problem in phonetics and given the fact that no two utterances are fully identical, how can neonates/infants possibly learn the phonemic contrasts/features and how are speakers to learn to categorize speech as required by their native language? More than that, how can one explain the ability of human beings to categorieze speech from a phylogenetic point of view? As for acquisition, the research findings on the perceptual abilities of neonates and infants before the onset of speech leave us with a paradox: Richly structured perception on the one hand vs. various imperfections in production including a wide range of deviations from target norms as well as large-scale intra- and interindividual variation on the other. Since infants can discriminate the sound contrasts of their native language(es) before the onset of speech their imperfections in production cannot be entirely due to perception. How, then, do perception and production relate to each other?

The role of perception within phonetics/phonology is such a huge topic that it needs to be broken down into individual issues, for example: (a) The impact of perception vs. production with respect to phonological units such as distinctive features, phonemes, and their typology. Are they rooted in perception or production? (b) The contribution of perception towards acquisition, either with respect to the morphophonology of lexical items or with respect to the inventory of phonemes. (c) The influence of perception and production on the evolution of sound systems and/or the human capacity for them. Can one be said to have taken the lead over the other? (d) The role of perception vs. production within the functioning of sound systems in everyday interactions, such as self-monitoring or coping with variation or style shifting. In fact, one may wonder whether there are any aspects that are not influenced at all by perception. I will primarily look at the first and the final issues, and I will touch on some of the others in passing.

2. DATA AND METHODOLOGY

For perception I rely on the literature; for production data on a project at the English Department of Kiel University, Germany, on 8 monolingual children acquiring German as their L1. Interviewing began before the children produced their first word. The data derive from spontaneous interactions recorded via a longitudinal one-hour weekly routine. The age range was 0;7-2;2. A wireless microphone was hooked to the child so that s/he was free to roam where s/he liked. The interviewer interacted with the child and the rest of the family in a normal way [8, 17]. The major purpose of the project was to provide data on the range of variation in early L1 production because of the paradox between perception and production noted above. It was thought that such data are required for any attempts to identify the cause(s) of the peculiarities in early L1 production. To capture the range of variation we recorded as many tokens per word per session as possible. The data do, in fact, reveal a large amount of inter- and intraindividual variation.

3. THE PERCEPTUAL BASIS OF SOUND SYSTEMS

3.1. Phonological units

One function of sound systems is to allow for the mapping of the variance inherent in the speech wave onto a finite set of categories. Units like phonemes and/or features serve this purpose. One task for the learner is to acquire these mappings. Given the segmentation problem of phonetics/phonology, this task presents a learnability problem. Acoustically the speech wave does not come readily segmented into distinctive features and phonemes. As a consequence, such units cannot be inferred, i.e. learned, from the input as available to learners unless the latter are pre-disposed as to what to look for. The respective ability must be innate to the organism. This conclusion is consistent with the research on neonates and young infants.

The crucial kind of evidence is where infants have been shown to be able to discriminate sound contrasts that do not occur in their ambient/native language. Such achievements cannot be due to learning in the sense of inferring information from sources external to the organism. At the present, there is a small but growing number of such studies. For example, Trehub [22] showed that English-environment infants aged 0;1-0;4 discriminated /a/ vs. /ã/ as in French and /ʒ/ vs. /ř/ as in Chzech. Streeter [21] found that 2-month old Kikuyu children in Kenya discriminated prevoiced /b/ vs. short-lag /p/ vs. long-lag /pʰ/, although their native language has only /b/ for the labials and prevoiced vs. short-lag for other places of articulation. Similarly Lasky et. al [13] reported that Spanish-speaking infants from Guatemala distinguished a long-lag category although Spanish does not have such a phoneme. Later Werker and her collaborators [23] found that English-environment infants were able to distinguish the dental vs. retroflex /t/ of Hindi and the glottalized velar vs. uvular stops of Thompson till around 0;7. More recently, several such studies have been produced by Best and her collaborators including clicks and ejectives [1, 2]; and there are also a few pertinent studies on vowels [11, 19, 20,]. (See [1,19] for more comprehensive overviews).

Although still scanty, data of this sort are highly suggestive with respect to such issues as the learnability, typology, ontogeny, and phylogeny of sound systems. Learnability is ensured because it is part of the structure of the auditory system to be particularly sensitive at certain points along a given acoustic dimension. Infants do not have to learn these points. They are part of their innate predisposition for handling natural human languages. As repeatedly suggested by Kuhl [9, 10, 11], these points of heightened auditory sensitivity can be thought of as constituting the biological basis of the distinctive features. They are simply the points at which auditory discrimination in humans is particularly sharp. Consequently, the typology of the distinctive features, i.e. the inventory of sound contrasts that natural human languages can chose from, is constrained by the points of heightened sensitivity.

Learnability and typology are interdependent as illustrated in Figure 1 on the basis of VOT in initial stops.

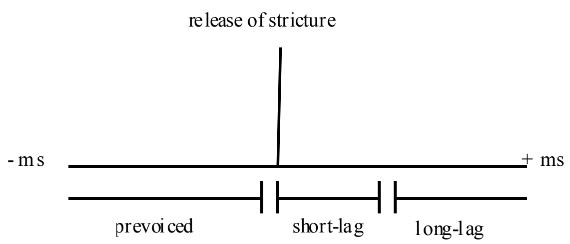

Figure 1: Typology of voicing distinctions.

Some languages have no distinction at all, e.g. Korean; many have two categories, like English, German, Spanish, or Kikuyu; the maximum number per language is three, as in Thai or Hindi. Typologically, languages may draw from the three options of pre-voicing, short-lag, and long-lag [15]. Research shows that adult speakers discriminate only the contrasts native to their language; young infants, however, distinguish the full set of the typological options with the category boundaries for adults and children at comparable points. This becomes evident if one adds up the evidence of the various infant studies where non-native sound contrasts were discriminated successfully, for instance, pre-voicing by anglophone infants [4, 18], long-lag by Spanish-speaking [13] or Kikuyu-speaking [21] children, and short-lag also by Kikuyu-speaking children. This evidence on VOT suggests that typology and acquisition relate to each other in such a way that neonates are born with the full set of typological options for the various acoustic dimensions. This enables infants to learn any language they happen to be born into. As they grow older their speech perception system is attuned to the ambient/native language. (For more details s. [25, 26]).

As for the evolution of sound systems, the kind of empirical evidence reviewed above for neonates and young infants, if linked with the research on speech perception in non-human animals, allows for an argument that retains the notion of innateness but dispenses with the assumption of language-specificity and/or species-specificity.

The points of heightened sensitivity in the human neonate are neither language-specific nor species-specific, because much the same sensitivities as with humans have also been found with chinchillas, macaques [9], and, more recently with some avian species [7]. Apparently, the auditory sensitivities of the human neonate do not constitute special speech detectors although they are indispensible for the structure of languages [ɑ,u]. This makes sense from the evolutionary point of view, because the parallels between humans and non-humans suggest that these auditory sensitivities are much older than *homo sapiens* and the human language capacity including the anatomy of the human vocal tract. If anything, the auditory sensitivities constitute the precursors to the language capacity and they were very likely also instrumental in the evolution of the quantal nature of the human vocal tract.

3.2. The status of the native language system

Whereas the points of heightened sensitivity can be regarded as innate functional properties of the auditory system, the categories of the native language of mature speakers are not innate but learned, namely, on the basis of the innate sensitivities. The former can be changed and/or unlearned over time; the latter probably cannot. The crucial evidence comes from Werker's research [24]. The experiment on the Hindi contrast of the dental vs. retroflex /t/ was run twice, the difference being the inter-stimulus interval. In one condition it was 1500ms, in the other 500ms. The subjects were monolinugual adult anglophones. They failed to discriminate the contrast in the 1500ms condition, but they were successful in the 500ms condition. This implies that the perceptual abilities present at birth remain available, but that they become difficult to access once the categories of the native language are established. At the present, it is not clear whether what learners rely on to develop their production is the acquired system of the native language perceptual categories, the original innate sensitivities, or both. (For more details s. [26]).

4. VARIATION IN PRODCUCTION

Given the auditory abilities of neonates and the perceptual development before the onset of speech, what may cause the variation in early L1 production including the deviations from target norms? No doubt, young children do have the ability to link perception to production at least by age 0;3-0;4 [12]. I suspect, therefore, that the peculiarties of early L1 production are to a large extent due to the maturation of the vocal organs and to the development of the (neuro-)motor abilities to control this apparatus and to coordinate the individual gestural components.

This is consistent with the findings from the project on L1 German mentioned above.

The rationale of the research was to start from the variation observed in fairly narrow transcriptions of the children's utterances. The variation, including the deviations from target norms, is summarized into a descriptive typology that lists the kind of deviations that can occur with a given stricture, e.g. what can go wrong with stops, glottals, consonantal clusters, etc. This typology constitutes the basis for determining the causes of the variation, i.e. whether they are due to perception, the nature of the mental representation, processing, or motor control. It is hoped that peeling away the imperfections due to lack of motor control and/or processing will help to reveal the structure of the representation of the lexical items as based on perception.

4.1. Syllable shapes

The syllable shape of a word may vary from token to token with respect to both onset and/or coda [14]. As for the latter, the data are not consistent with the view that children's first syllables are CV. For example, the tokens of the first 10 words recorded for child TL have CC, CVCC, CCVC, CCV, CVC, VCC, CV, and V. CC is short-lived; the others continue. Some of the codas agree with the target, others result from familiar processes like harmony or the deletion of final vowels in bisyllabic tagets; still others seem to reflect motor problems as when vowels are devoiced halfway through or when consonantal strictures are released too slowly, thus creating the impression of an affricate; and with some codas there is no discernible reason at all why a particular consonant should appear there.

Reduplication is not prevalent. In fact, the children do not even reduplicate consistently those tokens that reflect polysyllabic targets. As for the impact of perception, it does appear that the children are sensitive to the target right from the onset of speech. However, it would be premature, if one were to argue that the syllable as a phonological unit is derived from perception comparable to distinctive features and phonemes.

4.2. Discrepancies between perception and production

A sizable portion of early child phonological variation, it seems, is due to insufficient motor control, although the underlying representations of lexical items may already be much closer to the target. Insufficient stop gestures produce continuants; close misses of the alveolar place of articulation may result in dentals; etc. The typology of deviations is set up in such a way that it portrays the variation/deviations that occur for a given stricture/gestural mechanism. These deviations are subclassified as their sources, i.e. perception, representation, processing, or motor control. Consider glottal control, as involved in [h ʔ], syllable-initial vowels, aspiration, and voicing. Words that require any of these distinctions in the target are inspected as to how they are rendered by the child. By and large, the tokens of such words tend to show much the same pattern of variation in that all of the distinctions pertaining to the laryngeal mechanism may occur. Moreover, even words that do not involve this mechanism in the target may show the above pattern in those cases where the child does not produce the initial target consonant(s) [8]. As the children mature, the deviations are overcome according to a specific developmental sequence, with aspirated pre-voiced stops, for example, not fully controlled until age 6:0-7;0, as in Hindi [3].

Such deviations in production occur although the children can perceive the distinctions and will not confuse any of these words if addressed to them. This suggests that the children's representation of such words are much more advanced, if not target-like, although their production is not. What needs to develop is the motor control of the gestural mechanism rather than abstract phonological constructs. Moreover, there appear to be characteristic discrepancies between the development of perception and the (neuro-)motor control for the various gestural mechanisms.

4.3. Processing : Harmony

The evidence for harmony in the German children is much as reported in the literature. However, there is no child in our entire corpus who applies harmony consistently across-the-board in all lexical items that meet the appropriate prerequisites. In fact, not even all tokens of a given word are harmonized in the same way during the same recording session [8].

These observations confirm that the origin of harmony in children's speech is probably not in the mental representation of the word but somewhere within the processing that occurs in between the activation of the representation and the execution of the articulatory plan. This view agrees with the current idea that harmony is a device to ease the burden on production if one assumes that it is easier to produce the same stricture twice per word rather than two different ones.

4.4. Perception-based variation

German words like *da* (there) /dɑ/, *komm* (come) /kɔm/, or *Kiki* (a name) /kiki/ pose a special problem. Children may vary between [k] and [t] in all three words. In such cases harmony cannot be invoked as an explanation, because there is no velar stricture in *da* and no dental/alveolar one in *Kiki* or *komm*. Furthermore, generalization from words containing dentals/alveolars cannot be invoked either because such substitutions do not necessarily occur across-the-board. I am assuming that such deviations are due to the instanced-based way mental representations are created from the variability in the acoustic structure of the input a child is exposed to and how they are activated for production.

Recall that no two tokens of a word are acoustically identical. Thus Peterson and Barney's acoustic measurements of American English vowels [16] exemplified a considerable amount of overlap between phonemes. The data were recorded under favourable laboratory conditions. Only those instances were included in the analysis where the vowel was clearly identifiable as a good exemplar of the category. The overlap Peterson and Barney found was such that some tokens which were clearly identified by listeners as a specific vowel, say, /œ/ were located in central areas of adjacent vowels. If such is the range of variation in carefully elicited speech, one may suspect that the range of variation in real-life speech as inputted to children is at least the same, if not much larger. How is a child to decide which token to use for constructing his/her representation for a given lexical item or phonemic category, respectively? All learners can do is to go by what they encounter. That is, I am assuming that individual tokens are lumped together instance by instance. This will lead to central areas in the representation where the majority of tokens agree and to increasingly fuzzy fringes in those areas whére there is less agreement. The current hypothesis is that the kind of variation in early L1 speech as illustrated via *da, komm*, or *Kiki* above reflect the fuzzy nature of the representations reconstructed by the children on the basis of the input encountered by them. When activating a given representation there is no guaranty that children will always hit on the central part; in some cases the fringes may surface. (More details in [26]. See also [6]).

4.5. Articulatory patterns vs. segments

Although it is commonly assumed that phonological systems function on the basis of segments and phonological rules that apply across the entire lexicon this is not true for the early stages of phonological development. Initially, each word tends to have its own ideosyncratic pattern of variation suggesting that the underlying unit for coding phonological information is the individual lexical item [5, 26]. What first gives rise to a phonological system is a process that integrates the lexical items into articulatory patterns. They are subsequently broken down into smaller units. Both the structure of these patterns and their development varies considerably across children [17]. Apart from morphophonological reflexes of the target words the rise and the development of the patterns seems to be entirely a matter of the child's developing motor control.

5. CONCLUSION

Several points stand out. The first one is a word of warning. The data are still highly fragmentary. The range of non-native contrasts checked with young infants is anything but complete, and only very few contrasts have been tested on animals. In fact, there is Kuhl's report that monkeys failed to show the magnet effect the way human infants do [10]. Without replication studies we will not know whether this is a real difference or whether there was some methodological flaw. Nonetheless, the remaining evidence is consistent with the suggestions above. The second isssue is how to conceive of innateness. Is each point of heightened sensitivity innate to a given acoustic dimension, or do these sensitivities result from some general principle(s) of perception, as assumed by Kluender et al. [7]? Or maybe only some points of sensitivity are innate and others get established via some other procedure. The third issue is whether the points of heightened sensitivity are to be thought of as points on a given acoustic dimension or whether the effect derives from centers of gravity as in Kuhl's notion of a magnet [10, 11].

The final point relates to the perceptual-motor discrepancies noted above. There is a need for more detailed research on the development of the various articulatory mechanisms and how children learn to control them before we rush to highly abstract phonological explanations.

REFERENCES

[1] Best, C.T.1995. Learning to perceive the sound pattern of English. In Rovee-Collier, C. and Lipsitt, L.P. (eds.) *Advances in infancy Research* 9. Norwood: Ablex.

[2] Best, C.T, McRoberts, G.W. and Sithole, N.N. 1988. The phonological basis of perceptual loss for nonnative contrasts: Maintenance of dicrimination among Zulu clicks by English-speaking adults and infants. *Jounal of Experimental Psychology: Human Perception and Performance,* 14, 345-360.

[3] Davis, K. 1992. *Phonetic and phonological contrasts in the acquisition of voicing*. Ithaca: Cornell University.

[4] Eimas, P.D. 1975. Speech perception in early infancy. In Cohen, B. and Salapatek, P. (eds), *Infant Perception,* vol 2. New York: Academic Press.

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

[6] Jusczyk, P. W. 1997. *Disccovering spoken language*. Cambridge, MA: MIT Press.

[7] Kluender, K.R, Lotto, A.J., Holt., L.L. and Bloedel, S.L. i.pr. Role of experience for language-specificc functional mappings of vcowel sounds. *Mimeo*. University of Wisconson-Madison.

[8] Krüger, B. 1998. *Produktionsvariation im frühen Lauterwerb: Eine Typologie kindlicher Abweichungen von Modellwörtern*. PhD University of Kiel.

[9] Kuhl, P.K. 1987. The special mechanism debate in speech: Categorization tests on animals and infants. In Harnard, S. (ed.), *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

[10] Kuhl, P.K. 1992. Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories. In Tohkura, Y., Vatikiotis-Bateson, E. and Sagisaka, Y. (eds.), *Speech perception, production and linguistic structure.* Tokyo: IOS Press.

[11] Kuhl, P.K. 1995. Mechanisms of developmental change in speech and language. In Elenius, K. & Branderud, P. (eds*.), Proceedings of the 13th International Congress of Phonetic Sciences,* vol. 2, 132-139. Stockholm: KTH and Stockholm University.

[12] Kuhl, P.K. and Meltzoff, A.N. 1995. Vocal learning in infants: Development of perceptual-motor links for speech. In Elenius, K & Branderud, P. (eds*.), Proceedings of the 13th International Congress of Phonetic Sciences* vol. 1, 146-149. Stockholm: KTH and Stockholm University.

[13] Lasky, R.E., Syrdal-Lasky, A. and Klein, R.E. 1975. VOT discrimination by four to six-and-a-half-year-old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 215-225.

[14] Lindner, U. i. prep., *Silbenstrukturen im frühen L1- Lauterwerb.* PhD University of Kiel.

[15] Lisker, L. and Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustic measurements. *Word*, 20. 384-422.

[16] Peterson, G.E. and Barney, H.C. 1952. Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 3, 175-184.

[17] Piske, T. 1998, *Artikulatorische Muster und ihre Entwicklung im frühen L1-Lauterwerb*. PhD University of Kiel.

[18] Pisoni, D.B., Aslin, R.N., Perey, A.J. and Hennessy, B.L. 1982. Some effects of laboratory training on identification and discrimination of voicing contrast in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 297-314.

[19] Polka, L. and Bohn, O.-S. 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, 100, 577-592.

[20] Polka, L. and Werker, J. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.

[21] Streeter, L.A. 1976. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature* 259, 39-41.

[22] Trehub, S.E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47. 466-472.

[23] Werker, J. and Tees, R.C. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49-63.

[24] Werker, J. and Tees, R.C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Perception and Psychophysics,* 37, 35-44.

[25] Wode, H. 1994. Speech perception and the learnability of languages. *International Journal of Applied Linguistics,* 4, 143-168.

[26] Wode, H. 1997. Perception and production in learning to talk. In Hannahs, S.J. and Young-Scholten, M. (eds.), *Focus on phonological acquisition*. Amsterdam: John Benjamins.