Features, UG and 'Sensorimotor' Experiments on Infant Speech Perception

Mark Hale & Madelyn Kissock

Concordia University, Montréal

MFM26, May 2018

1 Overview

- Recent trends in infant (and adult) speech perception studies, especially in the psychological literature where much of the speech perception work is being and has been done, shows a growing focus on more integrated perception-production-sensorimotor (PPS) bases for perception (Werker & Gervain 2013). We look here at whether the results of such studies are significant for theoretical linguistics – specifically for the fundamental question of how the linguistic system is acquired. We examine a selection of recent experimental results, using Bruderer, Danielson, Kandhadai & Werker (2015) as the focal point (henceforth BDKW) from the PPS theory side.

2 Generative linguistic assumptions

- phonological representations consist of features
- the set of features is given by UG
- modularity – both of the linguistic system within the larger set of cognitive systems and of the domains internal to that system e.g., phonology vs. syntax
- competence-performance distinction
- Studies such as BDKW are of particular interest to our additional assumption that computation over linguistic (phonological) features is divorced from the properties of the interface system — articulatory-auditory — and is therefore 'substance-free' (à la Hale & Reiss 2008).

3 BDKW 2015

- Three experiments with 6 month old English environment infant subjects on a non-native contrast (dental [d̪] vs. retroflex [ɖ] voiced stop). (Note that only phonetic brackets are appropriate here although BDKW use, incorrectly, phonemic brackets.)
- Experiment 1 was a replication of earlier experiments on the above contrast and confirmed that, with only auditory stimulus, infants successfully discriminate this contrast.
- Experiment 2 and 3 employed two different types of teethers which were in the infants' mouths during the experiment. Experiment 2 used a broad, flat teether which lay on top of the tongue tip and blade, impeding movement. Experiment 3 used a curved teether that followed the upper/lower gum line and a which did not impede tongue tip/blade movement (labelled as the 'gummy' teether). Experiment 3 was intended to be a control for any distracting effects of simple presence vs. absence of a teether.

- All three experiments used looking time at alternating tokens vs. non-alternating tokens as the relevant variable, with longer looking time at alternating tokens indicating discrimination between those tokens.
- Reported findings of Experiment (2) (impeding teether) were that infants "...failed to show evidence of discriminating a phonetic contrast..." (BDKW p. 13533). In fact, this is an *average*.
- Reported findings of Exp. (3) were that infants "...successfully discriminated the Hindi /d̪/–/ɖ/ contrast, as did infants in experiment 1." (p. 13534)
- Overall conclusions of BDKW were:

  > "These findings implicate oral-motor movements as more significant to speech perception development and language acquisition than current theories would assume and point to the need for more research on the impact that restricted oral-motor movements may have on the delveopment of speech and language, both in clinical populations and in typically developing infants." (p. 13531)

- BDKW, while appropriately considering the possibility of a plus-minus teether effect, concluded that teething toys "...do not generally disrupt performance." However, upon the first exposure to alternating and non-alternating stimuli, the teether-free looking times differ by 2050 ms, while those with a teether show only 550 ms (Exp 2) and 350 ms (Exp 3) differences: it definitely looks like the teethers are significantly 'disrupting performance', perhaps not surprisingly distracting the child from attending to the phonetic contrasts. In the next trial, the child appears to have accustomed herself to the teether and now treats the phonetic contrast as more interesting (800 ms difference in both Exp 2 and 3), while the child who has already been attending the data seems to be less interested in the contrast now (500 ms difference, down from 2050 ms). [From our adult perspective, for example, the flat teether would be considerably more annoying!]

4 Lies, Damned Lies, and Statistics

- Longer looking time *differences* between the Alt and NAlt Pairs indicate *discrimination* (one should generally speaking not act in a different manner when exposed to the same stimulus). If BDKW are correct, we then predict that the *absolute* values (direction of differentiation is not relevant to whether discrimination is taking place, obviously) should pattern like this:

$$\textbf{No Teether} \cong \textbf{Gummy Teether} \quad > \quad \text{Flat Teether}$$
$$\textbf{Exp 1} \cong \textbf{Exp 3} \quad > \quad \text{Exp 2}$$

- Below are the results based on the graphs in BDKW (the values have a small degree of imprecision, therefore), and the patterning of the resultant values (grouping the two closest values with $=$ or $\cong$, and the most distant value with $>$, or, if the difference is very large, with $\gg$):

Run	Exp 1	Exp 3	Exp 2	Pattern		
Pair 1	2050	350	550	Exp 1	$\gg$	**Exp 2** $\cong$ **Exp 3**
Pair 2	500	800	800	**Exp 2** $=$ **Exp 3**	$>$	Exp 1
Pair 3	50	300	-250	**Exp 3** $\cong$ **Exp 2**	$>$	Exp 1
Pair 4	700	1175	-950	**Exp3** $\cong$ **Exp 2**	$>$	Exp 1
Average	825	656.25	37.5	**Exp 1** $\cong$ **Exp 3**	$>$	Exp 2

- In every case, the two teether experimental results (Exp 2 and 3) pattern more closely together than the 'no teether' experimental result. Only in the *average* of all pairs do we suddenly find Experiment 1 and Experiment 3 patterning together to the exclusion of Experiment 2. It is this averaging that the authors build their analysis around. But why does the average result diverge from that of every single pair?
- It is because, for reasons which are unclear *but which can have nothing to do with the ability to discriminate* in Experiment 2 the differential preference **shifts** in Pair 3 and Pair 4 to a *dispreference* (i.e., shorter looking time) for the Alt condition. Of course, the dispreference, like the preference, is only possible if one is distinguishing the two conditions. The negative values in the dispreference cases, when averaged with the positive values in the preference cases, yield an *average* value in which the consistent differentiation on Pairs 1, 2, 3, and 4 in Experiment 2 disappears! In fact, using the *absolute value* of the differences for Exp 2, we get an average magnitude of difference of 637.5, obviously directly comparable to the Exp 3 result of 656.25.

5 Conceptual Issues

- There is no evidence that infants at 6 months have sufficient motor skill development to produce the selection of sounds which they have been shown to discriminate, including the voiced retroflex token of this study. (Note that only articulatory gestures that are the result of deliberate commands to the relevant parts of the motor system can be included and not accidental, unplanned gestures that happened to produce a more articulatorily demanding sound.) Somewhat confusingly, BDKW state in their conclusions that "Sensorimotor information from the articulators selectively affects speech perception in 6-mo-old infants even without productive or perceptual experience with the speech sounds. These findings suggest that a link between the articulatory-motor and speech perception systems may be more direct than previously thought and is available even before infants accrue experience producing speech sounds themselves." (p. 13535). If the infants have not and *cannot* (due to immature motor skills) produce the required articulatory configuration, it is not clear what that sensorimotor feedback could consist of.
- The degradation of performance on discrimination tasks after experience with environment language (10-12 months, Werker & Tees 1984) cannot be attributed to loss of motor skills, which not only become more and more sophisticated at those ages but which are clearly retained in cases of bilingual L2 acquisition (Hale & Kissock 1997). It can only be attributed to the development of L1 featural representations (Hale & Kissock 2007).
- There seems to be a misunderstanding in the interpretation of causality in cited experiments such as Pulvermüller et al. (2006). Such studies suggest that perception of speech sounds causes activation in areas of the motor cortex, specifically, that there is a *link* between perception and the relevant motor areas for speech production. Such a link seems necessary to explain the acquirer's ability to reproduce a particular acoustic output via articulatory gestures with no instructions. However, since the perception event *causes* the activation of the motor areas, the motor activation itself cannot influence perception. In cases like Tio, Tiede and Ostry (2009) where the articulatory (in this case facial muscle) gesture was mechanically forced (and therefore presumably preceded or was simultaneous with the auditory stimulus) subjects performance with vowel-appropriate vs. vowel-inappropriate skin stretch only showed effects on the already-ambiguous vowel stimuli but not on clear tokens. Like the visual stimulus (McGurk-type) studies, and those involving the

lexicon (the Ganong effect), it is not surprising that information from other systems (in this case motor feedback) might be used to try to disambiguate the stimulus.

- While not proposed explicitly by BDKW, the only possible theory that might account for an actual causal effect of production-related events on perception is the 'forward prediction' model. However, Hickok (2012) offers a thorough discussion of studies of 'forward prediction' in perception based on transcranial magnetic stimulation (TMS) studies. He notes a number of problems with forward prediction and the studies that are used to support it. One of the most significant is that "...damage to the motor speech system does not cause corresponding deficits in speech perception as one would expect if motor prediction were critically important."(p. 399).
- The study ignores the long-known acoustic-articulatory inversion problem – the many-to-one relationship between articulation and acoustic signal. Lieberman & Blumstein (1988) pointed out early on that "... a human listener cannot tell whether a speaker produced a vowel like [e] by maneuvering his tongue...or by maneuvering his lips and larynx...unless the "listener" is equipped with X-ray vision or insists on holding conversations in front of X-ray machines." (p. 169). Dealing with the many-to-one mapping remains a challenge for speech recognition and other computer-based speech applications (Ji 2014).

6 General Discussion

- BDKW take as a foundation the assumption that, in adults, speech production 'impacts' speech perception, citing studies of McGurk-type effects from both visual and facial-sensory stimulus differences. However, such studies, including the cited Möttönen & Sihvonen (2005); Ito, Tiede & Ostry (2009); D'Ausilio et al (2009) and Möttönen & Watkins (2009) show only that an individual's behavior is the result of the aggregate effect of input from the many cognitive systems that humans possess – attention, memory, auditory, visual, conceptual, linguistic and so on. It is not at all surprising that the output of a single system may be masked by the combined output of all of the systems nor is it surprising that, any at particular point, the ratio of (apparent) contribution from one or another of these systems to the observed output will be different.
- As is often the case, there is a serious disconnect between statistically significant results and meaningful results. Nowhere is this clearer than in references by BDKW to the results of Kuhl et al. (2006) and Narayan, Werker & Beddor (2010), where BDKW summarize those studies as showing "...*improvement* [emph ours] in the discrimination of *native* [emph ours] speech sound contrasts." (p. 13531). The ability to discriminate a sound contrast is binary, not scalar — either an individual has the capacity (competence) to distinguish two inputs or they do not. The degraded performance of someone being repeatedly poked in the side with a stick does not indicate that their auditory *capacity* has also degraded.
- Related to the above, studies in this body of literature draw their conclusions from *averages* over subjects. Averages are of essentially no use to linguistic inquiry since we are interested in what the possible knowledge states of *individual* humans are. Not only is there no guarantee that an average represents *any* knowledge state, but, as in BDKW, conclusions about *individuals* ('language acquisition', e.g., is not a group enterprise) are being drawn from data that averages *across* individuals. The relevant data is not provided, but appears from BDKW's Figs. (4) and (5) that some subjects *did* discriminate with the flat teether. How is this possible, given their model?

- What would show that this is important for native language acquisition? Beyond the performance-based conclusions, there is no evidence that producing or not producing speech has any interesting effect on acquisition — no longitudinal evidence that children who have used flat teethers are delayed or prevented from acquiring a *grammar*.
- For linguistic purposes, the only significant results are those of Exp (1), a replication of the major contributions of Werker and others (Werker & Tees 1984 and many other such studies) to our knowledge of infants' ability to discriminate both native and non-native speech sounds from the youngest ages.
- Category-based discrimination (VOT), speaker-independent sound discrimination, and degraded discrimination performance with longer ISI's at 10-12 mos. on certain non-native sounds continue to provide support for a set of innate phonological features. That computation over those features is divorced from the physical reality of the interface systems remains a strong hypothesis.

7 Conclusions

- We argue that the role of audio-visual/motor/sensorimotor in speech perception is peripheral, and never causal, and that innate features (transduced from auditory input) are both necessary and sufficient for perceiving possible sound-based contrasts in natural language.
- In summary, we must assume that traditional experiments such as those favored by psychologists have a very different goal than subject-based data collection in linguistics. The former seek to describe the 'average', or 'majority' behavior (performance) of a selected population, rather than the cognitive systems and external conditions that come together to produce that behavior. Since, in real world performance, observed behavior is always a result of the confluence of effects of many systems, these are reasonable procedures for judging what kind of behaviors might exist. However, following the generative tradition of Chomsky and Halle (1968) and Chomsky (1957), our own (linguistic) interests lie in determining the properties only of the linguistic system — the types of of representations and computations possible in that module, along with some theory of how such knowledge comes to be instantiated in an individual (via a combination of UG and experience).

References

Bruderer, A., D. Danielson, P. Kandhadai & J. Werker. 2015. Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the USA*, vol. 112, no. 44, 13531-13536.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*. Cambridge: MIT Press.

D'Ausilio, A. et al. (2009) The motor somatotopy of speech perception. *Curr Biol* 19(5):381–385

Hale, M. & M. Kissock. 1997. Nonphonological triggers for renewed access to phonetic perception, in A. Sorace, C. Heycock & R. Shillcock (eds.) *Proceedings of the GALA '97 Conference on Language Acquisition,* University of Edinburgh, 229-234.

_____. 2007. Perception of non-native phonological contrasts: Evidence from and for featural representations, 15th Manchester Phonology Meeting, May 24-26.

Hale, M. & C. Reiss. 2008. *The Phonological Enterprise.* Oxford: Oxford University Press.

Hickok, G. 2012. The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model, *Journal Comm Dis* 45(6) 393-402

Ito T., M. Tiede & D. J. Ostry. 2009. Somatosensory function in speech perception. *PNAS* 106(4):1245–1248.

Ji, A. 2014. Speaker independent acoustic-to-articulatory inversion. Ph.D. dissertation Marquette University, Milwaukee, WI.

Kuhl, P. et al. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months, *Developmental Science* 9(2) F13-21.

Lieberman, P. & S. Blumstein. 1988. *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.

Sams, M., R. Möttönen & T. Sihvonen. 2005. Seeing and hearing others and oneself talk. *Cogn Brain Res* 2005 May (23:2-3): 429-35.

Möttönen, R. & K. E. Watkins. 2009. Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci* 29(31):9819–9825

Narayan, C., J. Werker & P. Beddor. 2010. The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination, *Developmental Science* 13(3) 407-420.

Pulvermüller, F. et al. 2006.Motor cortex maps articulatory features of speech sounds. *PNAS* 103(20):7865–7870.

Werker, J. & J. Gervain. 2013. Speech Perception in Infancy: A Foundation for Language Acquisition, *The Oxford Handbook of Dev Psych*, Zelazo P. D. (ed.) New York: OUP 909-925.

Werker, J. & R. C. Tees. 1984. Phonemic and Phonetic Factors in Adult Cross-Language Speech Perception, *JASA* 75(6) 1866-1878.