Loquens 6(2)
December 2019, e066
eISSN 2386-2637
https://doi.org/10.3989/loquens.2019.066

The perception-prodution link and linguistic theory

Mark Hale[1] & Madelyn Kissock[1]

[1]Concordia University, Montréal
mark.hale@concordia.ca, ORCID: https://orcid.org/0000-0002-9147-4513
madelyn.kissock@concordia.ca, ORCID: https://orcid.org/0000-0003-1736-6431

Submitted: 23/07/2019; Accepted: 23/07/2019; Published online: 12/01/2021

Citation / Cómo citar este artículo: Mark Hale & Madelyn Kissock (2019), The perception-production link and linguistic theory. *Loquens, 6*(2). e066, http://doi.org/10.3989/loquens.2019.066

ABSTRACT: 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.

Keywords: perception, production, acquisition, features.

RESUMEN: *El vínculo entre la producción y la percepción y la teoría lingüística.*– Las tendencias recientes en los estudios de percepción del habla infantil (y adulta), especialmente en la bibliografía de carácter psicológico en la que se ha enmarcado y se sigue enmarcando gran parte del trabajo sobre la percepción del habla, apuntan cada vez más hacia unas bases más integradas perceptivo-productivo-sensorimotoras (PPS) para la percepción (Werker & Gervain, 2013). En este trabajo analizamos si los resultados de tales estudios son significativos para la lingüística teórica, especialmente por lo que se refiere a la interrogante fundamental acerca de cómo se adquiere el sistema lingüístico. Examinamos una selección de resultados experimentales recientes, tomando como punto de referencia el trabajo de Bruderer, Danielson, Kandhadai & Werker (2015).

Palabras clave: percepción, producción, adquisición, rasgos.

1. BACKGROUND

There have been some interesting and, in some cases quite surprising, experimental results in the area of speech perception lately.[1] Our purpose here, given the brief space we have, is to explore one such study in a relatively detailed manner with a view to attempting to uncover the significance of the results for theoretical aspects of phonology and its acquisition. The study which we will focus our attention on is that of Bruderer, Danielson, Kandhadai and Werker (2015, henceforth BDKW). This study claims to have experimentally demonstrated that impeding the movement of the tongue-tip in 6-month-old English environment infants negatively impacts these infants' capacity to discriminate between a dental ([d̪]) and a retroflex ([ɖ]) voiced stop.

This result is surprising in several respects. First, it seems to indicate a relationship between productive capacity and perception which is counter-indicated by previous research on perception (Werker & Tees, 1984 *inter alia*). Second, it seems to necessarily locate the connection between perception and production in actual motor \ textit{activity} (as opposed to simply in, e.g., representations within the motor cortex). Of particular interest to us and to theories of phonology, these findings would indicate that our current understanding of the nature of features and how they relate to phonology is potentially flawed, especially the idea that phonology may be best conceived of as 'substance free'.

[1] The authors would like to thank an anonymous reviewer for their thoughtful and extensive comments and suggestions.

It is important to be clear about the assumptions which guide our work, particularly since they may not be shared by researchers in other fields, such as psychology. We base our analysis on the following assumptions of the generative phonology enterprise, as follows:

- phonological representations consist of (sets of) features;[2]
- the set of features is given by UG;[3]
- modularity – both of the linguistic system within the larger set of cognitive systems and of the domains internal to the linguistic system e.g., phonology vs. syntax;[4]
- one must distinguish between linguistic competence systems and the performance systems with which they interact;[5]
- Our final assumption is 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).

The paper is structured as follows. In Section 2, we provide a detailed account of the BDKW experiments along with a discussion of what results would be expected under BDKW's hypothesis as well as a discussion of the actual results. Section 3 presents several problematic issues in the BDKW study. Broader issues that arise in studies of this general type such as the relationship between the interface systems (articulatory-acoustic, in this case) and the linguistic computational system, as well as how data should be interpreted relative to each, are presented in Section 4. Section 5 discusses the role of 'influence' in assessing the cognitive capacities of the individual. We summarize the points presented in the paper in the final section.

2. BDKW 2015

2.1. The experiments

BDKW (2015) presents the results of three experiments designed to explore whether the well-established influence [in BDKW's view—mh&mk] of speech production on speech perception in adults is a result of experience or not [BDKW, 2015, p. 13531 abstract paraphrased].[6] To determine the answer to this, BDKW examine the relationship between speech production and speech perception in early acquirers.

The BDKW study first replicated earlier work on infant perception of non-native contrasts and then introduced an additional production-related variable to explore the primary question of a relation between production and perception.

The subjects in these experiments were 6-month-old English-environment infants. The infants were tested on the non-native contrast[7] between dental [d̪] and retroflex [ɖ] voiced stops.[8]

Experiment 1 was intended to replicate earlier experiments on infants' ability to perceive non-native contrasts, specifically here the contrast between voiced dental and retroflex stops. The auditory stimuli were tokens of [ɖ] and [d̪] culled from a single native speaker of Hindi. The experiment was set up in four trials for a total of 8 instances of 10-token sequences. In each trial, infants were presented with one each of the following – a 10 token sequence that was a repetition of the same token 10 times (Non-Alternating (NAlt) sequence) lasting 20 seconds and a 10 token sequence which randomly distributed 5 tokens of the dental segment, [d̪], with 5 tokens of the retroflex segment, [ɖ] (Alternating (Alt) sequence) also lasting 20 seconds.[9] There were four such trials per experiment (listed as Pair 1, Pair 2, Pair 3 and Pair 4 in the figures repeated from BDKW). As with comparable experiments, infant looking time was taken to be the measure of whether or not infants discriminated between differing tokens. Based on *general* trends in infant behavior, diverse stimuli (Alt condition) will attract infants' attention for a longer period than repeated identical stimuli (NAlt condition). Figure 1 of BDKW, repeated here, supports the conclusion that the infants note a difference between dental and retroflex voiced stops.

[2] `Features' means simply 'properties'. Entities which have all the same properties are necessarily the same entities — to differ, one needs to differ in at least one property. How such features are organized into groupings/sets is not of immediate relevance.

[3] That is, the set of features are part of the representational apparatus provided for the interface modules by the human genetic code. This follows by virtual conceptual necessity — properties of events in the world which the human genome does not provide any representational apparatus for can never be present in the mind, neither for short-term nor for long-term storage. As such, they cannot play any role in the development or functioning of a mental computational system such as phonology.

[4] Specifically, the relevant subset of the human representational apparatus made use of in linguistic computation differs from that used in extra-linguistic representation, and the set made use of in phonological computation differs from the set made use of in syntactic computation.

[5] That is, the properties of the linguistic computational system (its entities and processes) are distinct from the properties of those non-linguistic systems which feed it and make use of its outputs.

[6] The theoretical import of the term 'influence' here and elsewhere to describe the production-perception relationship will be examined in greater detail below, as it is critical to understanding the impact that such claims have on theories of the linguistic computational system.

[7] It is hard to explain the use of the term `contrast' in this context, given that we are discussing the perception of speech sounds by 6-month olds. The term is typically reserved for contrasting *phonological representations* that can only be deduced after a lexicon has been acquired (under standard assumptions, well after 6 mos.)

[8] BDKW fail to note the following relevant phonetic details: (1) that dental [d̪] is potentially present in English-speaking environments in words such as 'width' [wɪd̪θ]; and (2) the degree to which the realization of the alveolar stop in 'birdie' or 'hardy' may align more closely, in articulatory/phonetic terms, with their retroflex tokens in these phonetic environments than with 'normal' English alveolars. In addition, we note that only phonetic brackets are appropriate here although BDKW use, incorrectly, phonemic brackets when discussing these tokens throughout their paper. When coupled with the use of 'non-native contrast' discussed in the previous footnote, this engenders unnecessary confusion.

[9] An anonymous reviewer pointed out that 'random' is potentially significantly different than 'alternating' since it could result in 5 identical (to one another) tokens, e.g., 5 [ɖ]'s being followed by 5 identical (to one another) tokens, e.g., 5 [d̪]'s. Since the reported infants' average looking time is 10 seconds, the stimuli (and arguably therefore behavior) between Alt and NAlt could have been effectively non-distinct in some particular set of 10 tokens.

Experiments 2 and 3 were identical to Experiment 1 in every way except one. In each of these two experiments, the infant had a teether (held by the caregiver) in his/her mouth while hearing the stimuli.

Experiment 2 used a broad, flat teether which lay on top of the tongue tip and blade, impeding the movement of those parts of the tongue (labelled by BDKW as the 'flat' teether). Experiment 2 was the critical experiment used to explore any effect of production on perception by preventing the articulatory movements required to make the sounds in the auditory stimuli. Experiment 3 used a curved teether that followed the upper and lower gum line, which did not impede movement of the tongue tip or blade (labelled by BDKW as the 'gummy' teether). Experiment 3 was intended to be a control for any distracting effects of the simple presence vs. absence of a teether of any type.

Across the four trials there was a steady decline in average looking times, seemingly due to a 'familiarization to task' effect. As is the case in all such experimental designs, any significant difference in looking time between the two types of token pairings (in any trial) is taken to indicate that the child must be perceiving the phonetic difference present in the alternating tokens (otherwise, they would be perceived by the child as being identical to non-alternating tokens). The results of these three experiments are presented in BDKW Figure 4, repeated below as Figures 1–3.

BDKW summarize their results for Experiment 2 – with the impeding 'flat' teether – as showing that infants "… failed to show evidence of discriminating a phonetic contrast…"; their reported findings for Experiment 3 – with the non-impeding, gummy teether – were that infants "… successfully discriminated the Hindi /d̪/–/ɖ/ contrast, as did infants in Experiment 1" (pp. 13533–4).

The overall conclusion of BDKW was that "[t]hese 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 development of speech and language, both in clinical populations and in typically developing infants" (2015, p. 13531).

As described above, these experiments are designed to explore a potential relationship between production and perception where the former influences the latter. If such an influence exists, then we expect to see that in the form of certain experimental results. We describe the expected results below first, for purposes of comparison with the actual results.

2.2. Expected results

As we noted above, 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 of the looking time differences should pattern like this (direction of differentiation is not relevant to whether discrimination is taking place, obviously):

No Teether ≅ **Gummy Teether** ≠ Flat Teether
Expt. 1 ≅ **Expt. 2** ≠ Expt. 3

Figure 1: Average response time, all 4 trials, Experiment 1 (A).

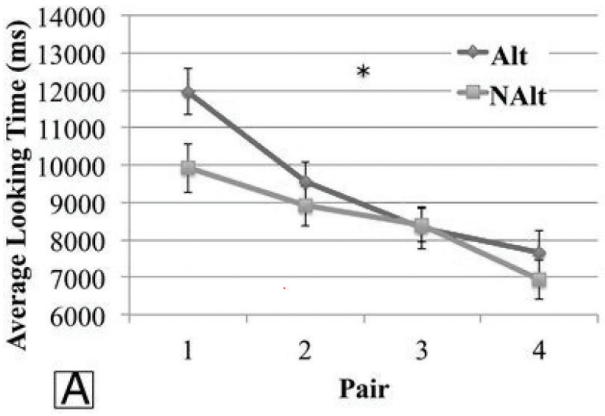


Figure 2: Average response time, all 4 trials, Experiment 2 (B).

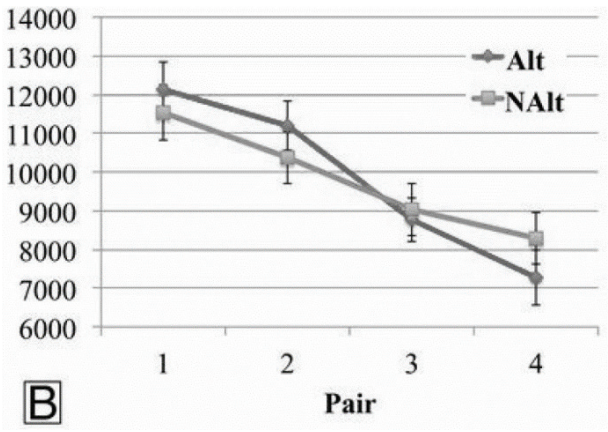
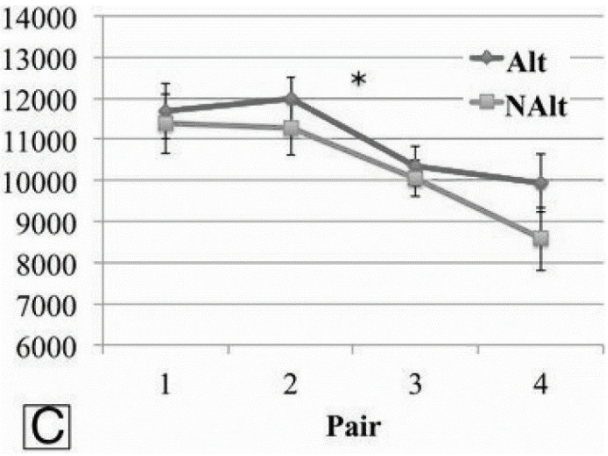


Figure 3: Average response time, all 4 trials, Experiment 3 (C).

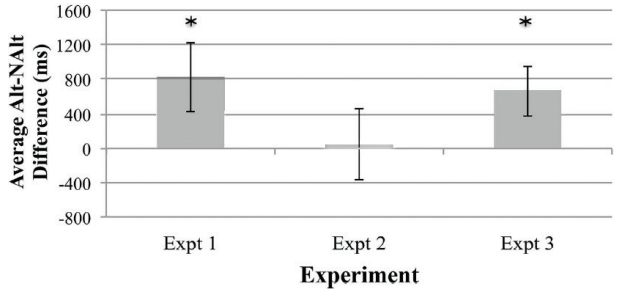


That is, we would expect average looking time differences for trials having alternating and non-alternating pairs (as all 4 trials did) with no teether (Experiment 1) and those with a 'gummy' teether (Experiment 3) to be essentially the same (since in these two conditions, according to BDKW, the child is perceiving the [d̪]–[ɖ] distinction) and to differ from the differences observed for those trials in Exp. 2 with a 'flat' (impeding) teether, as in Experiment 2 (since, according to BDKW, under these conditions the distinction is not being perceived, and the child is simply being presented with what is for him the 'same' data in the alternating as in the non-alternating pairs). BDKW predict, in particular, that the looking time differences between Alt and NAlt pairs will be greater in Experiments 1 and 3 than in Experiment 2. We turn now to the actual results.

2.3. Actual results

BDKW present the results for the average looking time, averaged for all subjects across all trials, in each of the three experimental conditions. In calculating differences in looking time between the alternating and non-alternating conditions BDKW took the Alt condition as a baseline, calculating the difference between the two conditions by subtracting the NAlt looking time from the Alt looking time.[10] This is summarized in their Figure 5, repeated below (along with the caption of BDKW) as Figure 4.

Figure 4: Alt–NAlt difference score averages. Average difference in looking time between Alt and NAlt trials for each experiment (in ms). Scores greater than zero indicate an overall Alt > NAlt preference. Error bars denote SEM difference, and an asterisk indicates a significant difference (from zero), as reflected in the individual ANOVAs.



This figure would seem to indicate quite unambiguously that there was discrimination between the [d̪] and [ɖ] segments in Experiment 1 and Experiment 3, but not in Experiment 2, as the authors concluded.

Below we present our attempt to render the results numerically, based on the graphs in BDKW repeated in Figure 1 earlier,[11] as well as the patterning of the resultant values (grouping the two closest values with = or ≅, and the most distant value with >, or, if the difference is very large, with ≫). Since discrimination is a function of the *magnitude* of the contrasting behavior, not on its *directionality*, we have used the *absolute values* of the looking time differences in establishing the 'pattern' column.[12]

Trial	Expt.1	Expt.3	Expt.2	Pattern
1	2050	350	550	**Expt.1** ≫**Expt.** 2 ≅ Expt.3
2	500	800	800	**Expt.2** = **Expt.** 3 > Expt.1
3	50	300	-250	**Expt.3** ≅ **Expt.** 2 > Expt.1
4	700	1175	-950	**Expt.3** ≅ **Expt.** 2 > Expt.1
Avg	825	656.25	37.5	**Expt.1** ≅ **Expt.** 3 > Expt.2

BDKW, while appropriately considering the possibility of a 'presence of teether' effect (this was the purpose of running Experiment 3), concluded that teething toys "… do not generally disrupt performance." However, upon the first exposure to alternating and non-alternating stimuli (i.e., in Trial 1 for each experiment), the no-teether looking times of Experiment 1 differ by approx. 2050 ms, while the with-teether times show only approx. 550 ms (Experiment 2) and approx. 350 ms (Experiment 3) differences.[13] This large difference between no-teether and teether trials strongly suggests that the teethers were significantly 'disrupting performance', perhaps not surprisingly distracting the infants from attending to the phonetic contrasts. In Trial 2, infants appear to have accustomed themselves to the teether and now treat the phonetic contrast as more interesting (800 ms difference in both Experiments 2 and 3), while infants who have already been attending to the data seem to be habituated to the contrasting tokens (500 ms difference, down from 2050 ms in Experiment 1).

In every trial, the two with-teether experimental results (Experiments 2 and 3) pattern more closely together than either of them does with 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 turns out that the average diverges 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*

[10] Note that this leads to negative values should the infants look longer at the NAlt condition.

[11] As noted above, the values thus have a small degree of imprecision.

[12] Therefore, for example, in Trial 3 the absolute value of the looking time difference in Expt. 2 is 300ms, and that of Expt. 3 is 250ms, making them significantly closer to one another than the 50ms of Expt. 1.

[13] All numbers are approximate because they had to be extracted from the graphic reprinted above – the paper provides no detailed statistics for the trials. Since BDKW found statistical significance in their numbers and our numbers are similar (but not actual), we can assume, but not assert, that our numbers also show statistical significance.

(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 by BDKW, yield an *average* value in which the regular differentiation on Trials 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 637.5ms, directly comparable to the Exp 3 result of 656.25ms.

Given that Experiment 3's average difference of approximately 656.25ms does not differ in a statistically significant way from Experiment 1's average of 825ms (this is central to BDKW's claim that the 'gummy' teether does not impede 'perception' of the relevant contrast), it seems pretty clear that the Experiment 2 average – taken now from the *absolute value* of the looking time difference – of 637.5ms will not do so, either. It certainly cannot differ in a statistically significant way from the 'gummy' teether result of 656.25ms. Impressionistically, it looks like there is an initial 'distraction' effect of having any kind of teether being held in one's mouth at all, which fades as one accustoms him or herself to its presence. The reason for the inversion in looking times for Trial 4 in Experiment 2 is not clear to us, but of course the entire experimental paradigm (difference in behavior entails perceptual distinction) entails that this result cannot be due to any failure to discriminate the alternating tokens. It should be noted that both familiarity and novelty effects have been found in the same experiment. As noted in McMurray and Aslin (2005, B20) "Although there are a number of hypotheses as to what factors affect the direction of preferences (i.e. novelty or familiarity), no consensus has emerged, and few studies make a priori predictions..."

3. STUDY-SPECIFIC ISSUES

We have noted that the actual results of these BDKW trials diverge from the expected results under their hypothesis that sensorimotor information may influence perception. Now we turn to more fundamental questions regarding the premises for such a hypothesis.

3.1. Articulatory Considerations

In the first case, we ask *how* the BDKW hypothesis could be correct – i.e., what the nature of speech perception would need to be for BDKW to get the results they believe they got in these experiments. Since impeding the actual articulation of the two stop segments being presented in the alternating pair condition would seem to be the relevant factor leading to the infants failure to discriminate, it seems we must conclude (as BDKW seem to do) that it is necessary to *articulate* a segment with a greater degree of accuracy than the 'flat' teether allows with respect to [d̪] and [ɖ] in order to perceive the two phonetic entities as distinct. The 'gummy' teether does not impede articulation in the relevant ways, and thus (the claim is) perceptual distinctness is maintained (as in the 'no teether' case). The teether, being in the mouth rather than the brain, can presumably only impede actual articulation, rather than, e.g., the motor cortex's *representation* of an articulatory target. Such a target must come into being before the articulation it gives rise to. Therefore, if the infant is to attempt the actual articulation of something like [d̪] and [ɖ] (and she must, or how could impeding that specific articulation lead to non-discrimination?), she must *first* build the relevant articulatory representation within the motor cortex. When no teether, or no 'impeding' teether, is present in the infant's mouth, the infant can implement in some way this representation. With the teether present, such an attempt to implement fails because of the impeding teether.

Given the above, it seems necessary to assume that infants can construct representations in their mind which can provide the basis for articulatory action. One possibility is that the same set of phonological features which capture the linguistically-relevant auditory properties of human speech are used to generate instructions to the motor cortex as well (a matter to which we will return below). However, as far as we know, there is no evidence that infants at 6 months have sufficient motor skill development to willfully and intentionally produce the selection of sounds which they have been shown to discriminate, including, for example, the voiced retroflex token of this study.[14] Indeed, the evidence regarding the speech production capacity of young infants indicates pretty clearly that their production systems are, from the perspective of adult production, seriously impaired. 'Seriously impaired' in this context means simply that, even when acting on the basis of relatively accurate articulatory representations (e.g., in the motor cortex), the infant cannot control and coordinate their systems well enough to produce adult-like, accurate-to-target articulations. If accurate perception of a [ɖ] requires that the child willfully and with some reasonable degree of accuracy implement the articulatory instructions for a [ɖ], we would expect the speech perception of a 6-month-old infant with no *additional* impediment beyond his own naturally impaired system to be exceedingly inaccurate. And yet perceptual studies, some of it by authors involved in BDKW's paper, have repeatedly shown that this is not the case.[15]

[14] If evidence of such motor skill exists, it must be included as foundational support for the study. Note that only articulatory gestures that are the result of deliberate commands to the relevant parts of the motor system could possibly explain the experimental results – accidental, unplanned gestures that happened to produce a more articulatorily demanding sound are not relevant in this context.

[15] It is worth pointing out that the well-known degradation of performance on discrimination tasks after experience with the environment language (10–12 months, Werker & Tees, 1984) obviously 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) in tandem with the `shedding' from that particular grammar of never-encountered features (Hale & Kissock, forthcoming).

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." (2015, p.13535). If the infants have not and *cannot* (due to immature motor skills) produce the required articulatory configuration, it is not clear what the 'sensorimotor information' referenced by BDKW could possibly consist of.

3.2. Casuality

A more serious problem with their hypothesis arises when attempting to determine what could cause the infant to construct the articulatory plan for a [d̪] or a [ɖ] when exposed to the alternating pair data. Clearly, the infant could only construct *distinct* representations for the two stops (which, according to the theory, they must do if they are to be able to perceive the distinction at all) if their brains have the information that the infant is being exposed to two distinct segments – and, to some reasonable degree, to the nature of the distinction between those segments. That is, the brain must accurately process and represent the distinction between [d̪] and [ɖ] in order to eventually find that it cannot physically implement the difference articulatorily. This capacity to process and represent the acoustics of speech seems definitional of 'speech perception', but BDKW clearly must mean something else by their use of this phrase, since their analysis could not logically stand otherwise.

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' (used most typically to describe the predictions of the motor system about expected sensory feedback from particular actions) in perception based on transcranial magnetic stimulation (TMS) studies. He notes both conceptual and empirical problems with forward prediction and the studies that are used to support it. These include: 1) divergent goals of forward prediction in motor as opposed to perceptual contexts – error vs. enhanced recognition, respectively; and 2) empirical evidence that shows that "...motor prediction in general tends to lead to a *decrease* [emphasis ours] in perceptual response..." (Hickok, 2012, p. 399). Most critically, however, Hickok notes 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" (Hickok, 2012, p. 399).

Finally, BDKW's 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" (169). Dealing with the many-to-one mapping remains a challenge for speech recognition and other computer-based speech applications as noted in Ji (2014).

4. BROADER ISSUES

The logical conundrum that requires the infants studied by BDKW to first discriminate perceptually between [d̪] and [ɖ] before they (eventually) can seem to fail to do so is paired with what appear to be insurmountable physiological barriers to obtaining results that would corroborate their hypothesis. Additionally, the broader considerations that motivated BDKW, such as results from adult studies and suggested links between articulation and perception, appear to be somewhat at odds with the BDKW study. We discuss these in turn below.

4.1. Lack of parallelism in adult studies

Studies with adult subjects differ in significant ways from the BDKW study.[16] Since BDKW state that "[t]he influence of speech production on speech perception is well established in adults" (BDKW, 2015, p. 13531) and are inquiring into whether the same effects are found in infants (as opposed to being the result of many years of experience), it is worth exploring these differences. First, a number of studies cited by BDKW differ crucially in the nature of the stimuli the experimental subjects were exposed to, and thus to the subjects' task in the experimental setting. One of the studies that was closest in its focus on articulatory gestures and perception was Ito et al. (2009). In that study, the articulatory (in this case facial muscle) gesture was mechanically stretched into positions that matched or did not match (one of) the articulatory gestures of the relevant vowels found in a continuum between the words 'head' and 'had' (minimally different in vowel height) – a stretching action that presumably preceded or was simultaneous with the acoustic stimulus although the timing was not noted specifically.[17] Ito et al. (2009) found an effect on subject performance (word identification) where the facial stretch influenced the subject's choice *in cases of ambiguous tokens only*, i.e., on the subset of stimuli which occupied acoustic positions between tokens of clear natural vowel qualities. Clear tokens were unaffected by facial stretch.

The Ito et al. (2009) stimuli (and, as a consequence, results) were critically different from the BDKW type. The tokens in BDKW were, we assume, intended to be unambiguous tokens of dental [d̪] and retroflex [ɖ]

[16] There are obvious reasons for certain differences given the difference in subjects -- infants vs. adults. These do not alter, as far as we can tell, the validity of the points discussed here.

[17] Trials with a non-speech-like stretching gesture were also included but had no significant effect on subject performance.

and were clearly distinguished by subjects in the no-teether trials. (The Ito et al.\ (2009) subjects were given ambiguous tokens in both types of trials – without facial stretch and with facial stretch – with poor identification of the ambiguous tokens in all types of trials.) The correct comparanda between the adult (Ito et al.) and infant (BDKW) experiments therefore would actually be the adult performance on *clear* tokens with (all of) the infant performance (all of whose tokens were clear). Adult performance in the clear cases was unimpeded by facial stretch according to the graphs presented in Ito et al. (2009) as well as their discussion of results. We have, as a consequence, conflicting rather than corroborating results in the Ito et al. (2009) study and the BDKW study (as reported by BDKW).

Ito et al. (2009) and similar studies reveal the hardly suprising fact that information from other systems (in this case motor feedback) might sway subjects toward a particular choice when faced with stimuli that are difficult to identify because the acoustic input is ambiguous. Any available additional source of information can and may be used by a subject to help disambiguate unclear acoustic input, including additional information from the lexicon (as 'Ganong Effect' studies demonstrate, Ganong 1980). The principle – using all available input to disambiguate confusing signals – is the same in visual-based studies such as the 'McGurk Effect' type (McGurk & McDonald 1976 and many subsequent studies). Here, however, there is a deliberate attempt to sway the subject by presenting conflicting acoustic and visual input and then forcing the subject to decide which, if any, of the input is an accurate representation of the incoming data. (A frequent result here is a 'combination' effect where subjects hedge their bets and use information from both inputs with a final decision locating the sound somewhere in between the two.) As in the Ito et al. (2009) case, the McGurk-type studies are not parallel to the BDKW study, though they differ in which aspect is not parallel. Unlike the McGurk-type studies, BDKW does not present conflicting external input to the subject. Instead, BDKW in some ways combine the physiological 'impediment' of the Ito et al. (2009) type with the clear acoustic input of the McGurk type, with the result that BDKW replicate neither type of study. What *is* common to all the studies just cited, however, is the multiplicity of cognitive modules being manipulated and tested. We return to the critical issue of identifying the multiple systems engaged in performance tasks in a later section.

4.2. Links between audition and articulation

The question of the nature of the 'link' between the articulatory-motor and speech perception systems is an important one. Studies such as Pulvermüller et al. (2006), cited by BDKW, demonstrate that perception of speech sounds causes activation in areas of the motor cortex, specifically, that there is a *unidirectional link* between perception and the relevant motor areas for speech production. Such a link seems conceptually necessary in any case to explain the acquirer's ability to reproduce a particular acoustic output via articulatory gestures with no overt instruction.[18] However, the evidence for the 'bidirectional relationship' asserted in BDKW does not seem to be equally well-supported, at least in terms of the linguistic module. The studies cited in support of a production-influenced perception suffer variously from problems discussed earlier, some of the McGurk type (Sams et al., 2005), some of the Ito et al.\ type (Möttönen & Watkins, 2009, which drew all its data from ambiguous/unclear tokens), and some of the TMS-based type discussed specifically in Hickok 2012 (D'Ausilio et al. 2009). As pointed out in a quote from Hickok (2012) earlier, if production were critical to perception, it would predict that damage to the motor speech system would result in perceptual failure – no such effect has been found.

What could be a possible 'link' between the articulatory-motor and speech perception systems in an infant which *pre-exists* articulatory experience, as alluded to in the quote above from BDKW's conclusions (p. 13535)? The mental structures and representational capacities of this type – which exist and have their properties prior to experience – are precisely the types of objects posited as the components of genetic UG by linguistics for some fifty years. In the phonological domain, this is generally taken to include a feature system used in the construction of lexical representations, and computed over by the phonological computation system. There was some debate in the early period of modern linguistics as to whether it was preferable to have these features represent *articulatory* properties of the segments found in human linguistic systems or the *acoustic* properties of those segments. One might have thought that the issue would have been resolved long ago – the facts of phonological computation could have easily favored one type of feature (e.g., the acoustic) as providing much cleaner analyses than those predicted by the use of articulatory features. Instead, the matter has largely been left to one side: Slavic linguists and those strongly influenced by the Jakobsonian tradition generally use acoustic features, while mainstream generative phonology and other intellectual descendants of SPE use predominately articulatory features.

In Hale & Reiss (2008) we find a proposal for why this might be the case: if phonological computation proceeds without regard for the alleged *substance* (articulatory *or* acoustic) of phonological features, it makes sense that it is difficult for scholars examining phonological processes to find unambiguous evidence in favor of one type or another. Hale & Reiss argue that the question is a non-issue: one and the same set of features is used by both interface systems – both the articulatory and the perceptual interfaces. It is the system-specific (motor vs.

[18] The capacity for linking an acoustic input to a motor plan that might produce such an acoustic effect seems to exist independently from the linguistic system, since non-speech sounds can also be reproduced. The linguistic system may well have taken advantage of a pre-existing ability.

auditory) transduction of this set of features that allows multiple system use. Both these interface systems being post-linguistic, linguistic computation has no access to the non-linguistic side of such transduction.[19]

The features of the phonological system – innate, and thus pre-existing any specific linguistic experience – provide just the desired 'link' between these two domains[20] of a six month old infant – as Werker and Tees (1984) firmly established – can construct a mental representation of the speech sound [ɖ], presumably leveraging this pre-existing, innate feature system and the (likewise innate) transduction system(s) for perception. This same feature representation can be transduced to an articulatory plan which, however, *with or without an impeding teether* the 6-month old's performance systems will be incapable of accurately implementing.

Under such a conception of matters, the claims made by BDKW on the basis of their experimental data cannot be true. We have detailed several reasons for this: (1) the infant must perceive the distinction between [d̪] and [ɖ] in order to *attempt* to articulate it and subsequently be impeded in so doing by a teether; (2) the unified featural representation made use of in both perception and production is involved in the transduction of either modality using innate mechanisms; and (3) as we have just argued, the 6 month old infant's performance systems fail to accurately implement articulatory targets regardless of the presence or absence of artificially added impediments – yet the perception of phonetic contrasts at this stage is quite robust.

BDKW use a scale which somehow counts looking for 300ms *longer* at the alternating condition as indicative of discrimination, but looking for 300ms *shorter* at that same alternating condition as *failure* to discriminate. Fortunately, a simple reworking of the statistics heeding the *magnitude* (i.e., absolute value) of differentiating looking times for the alternating vs. non-alternating conditions yields sensible results. What appears to actually be going on in BDKW's experiments is that the presence of the teethers, being held in place in the mouth of the infant by a caregiver for the duration of the testing, triggers a modest distracting effect with respect to *attention* to the segmental contrast present in the alternating trials, which fades on subsequent trials with a teether. The decline arising from task familiarization shows a similar trajectory both without any teether and with both kinds of teethers after that point. The discrimination-evincing behavior in the 'impeding' teether experiment shows an interesting – and at present unexplained – shift in looking preference for the NAlt condition over the Alt condition. If this is more than a statistical fluke, further work should be pursued to try to clarify just what it represents: but what is clear is that it cannot be explained by BDKW's claim that there is a 'failure to discriminate' the alternating stimuli pairs in this trial. As Houston-Price & Nakai (2004, p. 344) note regarding the use of this type of 'looking preference' experimental design: "'[t]he direction of a looking preference is largely irrelevant when infants' discrimination ability or recognition memory is of primary interest; any deviation from random behavior indicates that a difference between the stimuli has been detected.''

It is worth noting, as Houston-Price & Nakai go on to say (2004, p. 355), that "some infants will progress through the sequence of preferences more rapidly than others; if data are averaged across a group of infants, looking preferences may appear to be random at certain time points, despite individual infants showing clear familiarity or novelty preferences…" Thus, while we have emphasized the dangers of averaging across trials, especially with respect to how one establishes the values being averaged, we also note that averaging over sets of subjects gives rise not only to the methodological concerns raised above, but to the problem that the object of linguistic research is the knowledge state of an individual. 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) appear to be being drawn without the direct consideration of the actual behavior of any individual subject. The relevant data is not provided, but the numerical data cited is fully consistent with *some subjects* (if a significant minority) in fact showing 'discrimination' even under BDKW's misanalyzed statistics in the flat teether experiment. But how could this be possible, given the explanation they propose?

5. KNOWLEDGE VS. BEHAVIOUR: THE ROLE OF 'INFLUENCE'

BDKW take as a foundation the assumption that, in adults, speech production 'influences' speech perception, citing studies of McGurk-type effects from both visual and facial-sensory stimulus differences, *inter alia*. However, as we pointed out above, such studies, including the cited Sams, Möttönen & Sihvonen (2005); Ito, Tiede & Ostry (2009); D'Ausilio et al. (2009) and Möttönen & Watkins (2009), show only that an individual's *behavior* when confronted by particularly challenging stimuli[21] 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 other systems, nor is it surprising that, any at particular point, the ratio

[19] Although this claim is confined to the linguistic system and its relationship with additional systems (e.g., acoustic and auditory) that interact with it, it seems likely that other (non-linguistic) auditory and visual input have some corresponding type of representation-interface systems.

[20] An explicit discussion of the link between, in our view, substance-free phonological features and the external systems which interact with them is somewhat orthogonal to the matter at hand as well as too complex to treat fully within our space limitations.

[21] For example, stimuli synthesized to be precisely half-way between an [i] and an [u], or stimuli presented in a context in which much of the subject's visual field is taken up by a large image of an articulating mouth, which the subject might reasonably conclude the experimenter would like him/her to attend to.

of (apparent) contribution from one or another of these systems to the observed behavioral output will be different. The way to clarify the role of each of these systems in giving rise to the observed data is to build as unambiguous and restrictive a model as is possible for each domain, rather than to weaken our model of each domain by allowing vague types of interpenetrability.

We see no evidence that the role of visual and sensorimotor systems in speech perception is anything other than peripheral – no causal relationship has been established – and that innate features (transduced from auditory input) provide the necessary representational apparatus for the extraction of phonetic features in a clearly presented acoustic speech stream. Once represented in the feature system, the representation is available for articulatory 'playback' via transduction by the production systems, though in the infant these systems are normally inadequate for the generation of adult-like articulatory acts.

It is reasonable to ask, however, whether such peripheral 'influences' could have a significant (i.e. long-term/permanent) effect on acquisition (by preventing the acquirer from accessing the linguistic portion of the data of the input). BDKW, in their conclusions, call for:

> … the reconsideration of theories concerning the processing of speech during language acquisition: Such theories must account for the influence of sensorimotor information on speech perception and determine the consequences or advantages of such a linkage as infants acquire the native language. (BDKW, 2015 p. 13535).

We asserted above that innate features provide the necessary apparatus for perception of speech sound contrasts as supported by BDKW and many earlier studies. However, as we discuss the role of influence, an important question arises of whether those innate features are *sufficient* for successful acquisition. That they are *not* sufficient is the only interpretation of the BDKW hypothesis that we can imagine. This would entail that, extra-linguistic, physiological factors put acquirers at risk in developing a linguistic system.[22] Could, as BDKW state, 'influences' of this type be in any way critical to development?

As far as we know, support for such a hypothesis is completely absent. Data showing a critical impact (of extra-linguistic, physiological factors) on language development would need to show that children are affected *in the long term* – i.e., not simply in the trials of an experimental setting but over the course of the years of acquisition, resulting in a *failure* to acquire some aspect(s) of the environment language. One of the clearest instances of counter-evidence to production-perception influences demonstrated in the long term) of the type BDKW are suggesting can be seen in cases of infants who have undergone tracheostomies and been intubated for lengthy periods. Hill & Singer (1990), for example, studied infants who underwent this type of procedure at a mean age of 4.2 mos. and had the endotracheal tube in place for a minimum of 3 mos. Their conclusion, after testing the children at a little over 5 years (mean age), was that only *expressive* capacity (if any) was affected by the long-term disruption of speech-related laryngeal activity (due to the endotracheal tube). They state:

> For the entire group of children, the overall measure of language functioning at follow-up were within normal limits and commensurate with cognitive ability. However, when a breakdown of results based on the children's ages was done, a clear pattern of language disability was noted in the *expressive* [emph. ours] language of the oldest group of children tested. (Hill & Singer, 1990, p. 15).

Vocal-fold vibration is inhibited by tracheostomy/intubation; however, such vibration is required in the production of all voiced speech sounds and voiced sounds make up the majority of sounds heard in human running speech. If perception were critically influenced by production, the inability to produce voicing should have had a significant deleterious effect on perception of voiced sounds, resulting in failure to acquire normal receptive language (i.e., a comprehension deficit). As the study indicated, no such effect was found.

Furthermore, although completely anecdotal, it is relevant to note neither the virtual barrage of confused information from the physical world surrounding the infant nor the infant's own non-speech oral activities (chewing, sucking, crying, making random noises), all of which occur routinely while getting language input from the environment, appear to have any interesting effect on acquisition. The vastly sub-optimal conditions under which language must be processed were pointed out by Chomsky (1965):

> [the competence model assumes that the speaker-listener] knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for modifying it has been offered. To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena. (Chomsky, 1965, pp. 3–4).

At this point, evidence points to the innate perceptual ability of infants as being not only a necessary but also a sufficient condition for the acquisition of the phonological representations critical to language.

[22] Assuming normal physiology and environment.

6. CONCLUSIONS

In this paper we have argued that BDKW's surprising claims about the role of articulation in speech perception arise from a statistical misinterpretation of significant aspects of their experimental results. We have also noted that the claims suffer from a logical flaw, requiring an incoherent conception of normal causality. In addition, we have called into question the validity of claims about the influence of non-linguistic stimulus on linguistic *knowledge*, noting that failure to attribute correctly the sources of performance/behavior to the appropriate cognitive systems impedes rather than improves our understanding of such systems.

The BDKW study was not selected by us at random. Leaving aside its statistical shortcomings, the study reveals the wide gulf that seems to us to exist between the goals of much experimental literature regarding the linguistic system and those of the contemporary theoretical linguist. The former seems to be seeking 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 on a particular occasion in any single individual. Since, in real world performance, observed behavior is always a result of a confluence of the 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 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).[23] A theory which posits 'substance-free' phonological features as a core component of the UG-provided knowledge the acquirer brings to his or her task seems to provide key insights into the pre-experiential 'link' between speech perception and speech production which BDKW express an interest in. Experimental work guided by explicit and clear theoretical assumptions (whether of the 'substance-free' type, or assuming some other underlying 'linkage') remains a desideratum. The innate capacity for a perceptual \ textit{and linguistic featural} system in humans seems to already be well-established by earlier experimental results (Werker & Tees, 1984 and Eimas et al.\ 1971 \ textit{inter alia}) and provides the most productive foundation for further research.

[23] As we noted earlier and as is well-established as a conundrum in the theoretical literature, the experience that the acquirer gets must be extrapolated from the effects of all of the external and internal extra-linguistic input. To date, however, all evidence points to the acquirer's success in this domain barring, of course, special physical or physiological situations which prevent input from being received.

7. REFERENCES

Bruderer, A., Danielson, D., Kandhadai, P. & Werker, J. (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. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

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

D'Ausilio, A., Pulvermüller, F, Salmas, P., & Bufalari, I. (2009). The motor somatotopy of speech perception. *Current Biology, 19*(5), 381–385. http://dx.doi.org/10.1016/j.cub.2009.01.017

Eimas, P., Siqueland, E. R., Jusczyk, P. and Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306. http://dx.doi.org/10.1126/science.171.3968.303

Ganong, W. F. III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110–125. http://dx.doi.org/10.1037/0096-1523.6.1.110

Hale, M. & Kissock, M. (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* (pp. 229–234). Edinburgh: University of Edinburgh.

Hale, M. & Kissock, M. (2007). Perception of non-native phonological contrasts: Evidence from and for featural representations, *15th Manchester Phonology Meeting*. Manchester. http://www.lel.ed.ac.uk/mfm/15mfm.html

Hale, M. & Kissock, M. (Forthcoming). The regularity of syntactic change. In T. Eythórsson and J. G. Jónsson (Eds.) *Syntactic Features and the Limits of Syntactic Change*. Oxford: Oxford University Press.

Hale, M. & Reiss, C. (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 of Communication Disorders, 45*(6), 393–402. http://dx.doi.org/10.1016/j.jcomdis.2012.06.004

Hill, B. & Singer, L. (1990). Speech and language development after infant tracheostomy. *The Journal of Speech and Hearing Disorders*, *55*(1), 15–20.

Houston-Price, C. & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341–348. http://dx.doi.org/10.1002/icd.364

Ito T., Tiede, M. & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, *106*(4), 1245–1248. http://dx.doi.org/10.1073/pnas.0810063106

Ji, A. (2014). *Speaker Independent Acoustic-to-articulatory Inversion*. Ph.D. Dissertation. Marquette University, Milwaukee, WI.

Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-21. http://dx.doi.org/10.1111/j.1467-7687.2006.00468.x

Lieberman, P. & Blumstein, S. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge: Cambridge University Press.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264* (5588), 746–748. http://dx.doi.org/10.1038/264746a0

McMurray, B. & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, *95*(2), B15-B26. http://dx.doi.org/10.1016/j.cognition.2004.07.005

Möttönen, R. & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of

speech sounds. *Journal of Neuroscience*, *29*(31), 9819–9825. http://dx.doi.org/10.1523/JNEUROSCI.6018-08.2009

Narayan, C., Werker, J. & Beddor, P. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science, 13*(3), 407–420. http://dx.doi.org/10.1111/j.1467-7687.2009.00898.x

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martín, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865–7870. http://dx.doi.org/10.1073/pnas.0509989103

Sams, M., Möttönen, R. & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23*(2–3), 429–35. http://dx.doi.org/10.1016/j.cogbrainres.2004.11.006

Werker, J. & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology* (pp. 909–925). New York: Oxford University Press.

Werker, J. & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of Acoustical Society of America*, *75*(6), 1866–1878. http://dx.doi.org/10.1121/1.390988