

1 The Emergence of Phonological Representation

PATRICIA DONEGAN

1. Introduction

Linguistic emergentism assumes that the properties of language arise from the interaction between the demands of communication and general human capabilities. In phonology, this means that phonological representation and processing arise for reasons that are largely phonetic. *Phonological representation* refers to the mental representation of lexical entries – to speech forms as remembered and intended. The intentions of speakers and the percepts of hearers are not like the generativists' *underlying* representations. Rather, phonological intentions and perceptions are *phonemic* (more like generative *lexical* representations). A central problem to be addressed here is the difference between the remembered, intended phonemic representations, and the phonetic representations targeted for actual pronunciation.

The emergence of phonological representation (phonemic or morphophonemic, see subsections 5.3 and 5.4) raises some potentially puzzling questions: Why do speakers form phonemic representations? How and when does phonemic representation “emerge” in acquisition? How do children learn which features they must attend to and which they may ignore? How do lexical forms “emerge” from the speech stream in perception?

The sections below will attempt to establish the following: 1) the automatic or “natural” phonology of a language, the system which underlies the perception and production of speech, is distinct from the morphophonology or “lexical phonology,” which is a system of conventionalized sound-alternations that are a result of the language's history; 2) phonological processes, the automatic responses to constraints on speakers' articulatory and perceptual abilities, are phonetically motivated mental substitutions; 3) phonemic perception and representation are based not on distribution, but on the interaction of phonological processes; 4) children's development toward adult-like perception of speech sounds relates perception to production, and perception can consequently be characterized as “phonemic” quite early – even before the child begins to speak; 5) adult speakers undo multiple processes in interpreting phonetic cues, to arrive at lexical forms.



2. Phonology Is Not Morphophonology

Phonological processes emerge from the requirements of speaking and hearing. Lexical or morphophonological rules, on the other hand, are conventional and emerge historically. The morphologically conditioned alternations that affect word formation, derivation, and inflection may distinguish among foreign and native word-stocks (lexical strata), morphological classes, etc., and they are always alternations of phonemes. But what makes a speaker sound like an English or French or Korean native speaker and what makes him perceive other languages, at least at first, in terms of his own system is the process phonology of his language.

2.1 *Morphophonology is conventional*

Morphophonological alternations, like the voiceless/voiced pairings of f/v, θ/ð, s/z, in alternations like *loaf* (n.)/*loaves*, *breath*/*breather*, *house*/*housing* may appear to refer to phonetic features, and indeed, this alternation arose as intervocalic voicing. But English now admits exceptions to such voicing, as in *chief*/*chiefs*, *loaf* (v.)/*loafing*, *mouse*/*mouser*/*mousing*, even with the same affixes that sometimes require it. Such morphologically conditioned alternations no longer respond to the phonetic (in)abilities of speakers; they are purely conventional. The rules that account for these alternations, which are not *synchronically* phonetically motivated, arise across generations (see Donegan and Nathan, in press; Johnsen, 2012). Their historical origins may of course be phonetic, but in the living language they are based on tradition. They may remain quite regular, and productive enough to apply to nonce forms – in the appropriate morphological environment. But they do not represent limitations on speaker abilities: *[dɪˈvaɪnɪti] (for *divine-ity*) and *[mɔɪstən] (for *moist-en*) are not unpronounceable in English – they are merely unacceptable.

2.2 *Phonological processes are “natural”*

Phonological processes account for the discrepancies between lexical representation or intention, and phonetic target. The feature changes that create these discrepancies are responses to innate limitations on perception and production, and they are consequently universal. But the phonologies of different languages differ, because each language requires that its speakers inhibit some substitutions, learning to pronounce their inputs. From language to language, process inhibitions differ. And, as noted in discussions of Optimality Theory, different languages may select alternative ways of avoiding a difficulty (e.g. final voiced stops are often devoiced, but some languages postnasalize them instead).

Automatic or natural processes, unlike conventional rules, apply in speech processing – not “in the lexicon.” They may create allophones or change phonemic interpretations. Together with the prosody of a language, they account for the native “accents” of speakers. They often apply across word boundaries. They affect perception and pronunciation of second languages (“interference”), and they underlie systematic variation and sound change.¹ They also create phonological representation.



Noting common phonological patterns that appear in children's speech and in cross-language comparisons, Stampe (1969, 1973) referred to automatic phonological processes as innate. This has been somewhat misconstrued: the processes are responses to the limitations of the human faculty for speech perception and production, but this need not mean they are part of a genetically endowed "language faculty" or "Universal Grammar."² Processes may be discovered by children in using the body (as in babbling) – or by adults in attempting to produce new, foreign sounds (like implosive or ejective consonants). It is the phonetic abilities of the speakers that are innate; the responses or processes may be discovered, automatically and unconsciously, when one attempts to use and expand one's articulatory abilities.

3. Processes Are Both Phonetic and Phonological

Some phonologists say that processes that refer only to phonetic qualities and not to morpheme boundaries or classes are phonetic. Processes do indeed have physical motivations (articulatory and perceptual), but processes are mental.

3.1 *Phonetic motivation, mental (phonological) application*

A *constraint*, as seen here, is a limitation on a speaker's phonetic abilities. A constraint against a particular simultaneous combination of features or against a particular sequence means that a speaker is unable to produce that configuration (or, in some cases, perceive it) because of some inherent phonetic difficulty. A *phonological process* changes a representation that presents a phonetic difficulty into one that lacks that difficulty (Stampe, 1973: 1). Because processes respond to phonetic causes, they are storable in terms of phonetic and prosodic information only. Children are subject to more constraints and apply more substitution processes than adults. Learners overcome constraints, suppressing the processes, by learning to perceive and pronounce the more difficult configurations.

The model presented here is derivational. In Optimality Theory, overcoming a well-formedness constraint is characterized as demotion, ranking the constraint below a particular Faithfulness constraint (or, perhaps, below an alternative constraint that rules out the difficulty). But the processes cited here compare only to the phonetically motivated constraints of OT, not to conventional or language-specific constraints.

Processes are not physical "slips" or accidental mistimings. They apply in the central planning of speech, adjusting the intention (the lexical representation) to create an altered target (the phonetic representation). We cannot expect the tongue or velum or laryngeal muscles to make consistent changes like nasalization, labiality spreading, loss of coronal closure, glottalization, etc. "on their own." Processes do not happen "in the mouth" – any more than the balancing movements of a bicyclist happen "in the arms" or "in the shoulders" (cf. Whalen, 1990; Moosmüller, 2007; Kingston and Diehl, 1994). And though universally motivated, processes may apply or not in different languages. For example, nasality assimilation makes Korean /kok-mul/ → [koŋmul] 'grain', or /kjəp+mun/ → [kjəmmun] 'double door', but this assimilation does not apply in English. Conditions on a process may also vary: stops assimilate to a following



non-coronal in both English and Korean, but in Korean both labials and coronals may assimilate to a following velar (Jun, 1996), whereas English speakers limit the input to coronals.

Because they are part of the mental processing of speech, and because their application varies from language to language, processes are *phonological*. Yet they reflect actual (in)abilities. Even a careful pronunciation like ['kʰǣnt 'pʰleɪ̃] fails to match the intended but unpronounceable / 'kænt 'pleɪ̃ /. And other pronunciations of *can't play*, though they may be quite ordinary, like ['kʰǣmp 'pʰleɪ̃] or ['kʰǣpʔ 'pʰleɪ̃] or ['kʰǣʔ 'pʰleɪ̃], diverge from the phonemic/lexical intention even more. The regular substitutions of adults that create these divergences (vowel nasalization, stop aspiration, coronal assimilation, glottalization, nasal deletion, etc.) are consistent patterns that a phonology must account for. Adults find it difficult or impossible to avoid these adjustments – just as a child may find it difficult or impossible to say the final velar of *dog* without assimilating the initial stop to its velarity, so that *dog* is [gɔg]. So, unlike morphophonological rules, processes affect production in adult second language.

3.2 *Features are associations of sound quality to gesture*

Unlike morphophonological rules, which always result in the substitution of phonemes, phonological processes apply in terms of features. If we take the phonetic motivations of phonological substitutions seriously, features are essentially linked to phonetics. Each feature has two phonetic aspects: articulatory and auditory. Since the form and capabilities of the human vocal tract are reasonably consistent, we find consistent (though not entirely invariant) relationships between articulatory movements and perceptual qualities; for example, lowering the velum has consistent (though not identical) acoustic effects across speakers.

A child who hits a tray with a spoon forms a link between the action and the resulting sound. Similarly, a child who vocalizes and babbles forms links between his gestures and the resulting sound qualities. Although a learner's earliest representations of first-recognized words may be "holistic," involving knowledge of only a general acoustic shape, like a dog's representation of the sound of *Sit!* or a horse's of *Whoa!*, important changes take place during the period of early vocalization and babbling. In establishing connections between gestures and auditory effects, a child learns what kinds of gestures produce particular sound qualities. The child's own vocalizations, observation of speech activity in others, and active articulatory practice (with auditory self-monitoring) all influence the development of this system of connections (cf. Fry, 1966; Locke and Pearson, 1992; MacNeilage, 2008, among others).

If one says [papapa], one hears an alternation of very low intensity and relatively low frequency with a slightly noisy burst and a much higher-intensity, non-noisy sound. This can be associated with a relatively closed jaw with lip closure, and change to a relatively open jaw. The same oral articulation, with a lowered velum, produces the acoustically different [māmāmā]. Children must make such articulatory-acoustic connections in order to imitate adult forms. These links or mappings between articulatory gestures and auditory or acoustic properties are *features*. They are the knowledge that underlies the ability to imitate, which requires linking the properties of sounds heard to articulations that produce them.



Auditory correspondents to particular articulations may be complex, and acoustic qualities may be associated with complex articulations, but consistent cues in a speaker's own utterances underlie these connections. As the child creates a global mapping of gestures to sensory outputs, variants of a movement occurring in different contexts and sensory input of multiple kinds (tactile, proprioceptive, visual, auditory) are integrated into the mapping. This entails considerable complexity, but no more than other mappings that are required in learning to use the body. And because features are complex articulatory-auditory connections, hearers can use a variety of acoustic cues to identify a speaker's articulatory targets and arrive at his lexical intentions.

4. Phonemic Perception and Representation

A "naive" adult speaker ordinarily does not notice or reproduce a phonetic difference between sounds unless that difference represents a phonemic difference in his or her language.

4.1 *Evidence for phonemic representation*

The phonemic perception and representation of speech is one of the best-established constructs of linguistic theory. The morphologically conditioned alternations of a language always reflect phonemic – never allophonic – changes (structure preservation). The widespread use and easy learnability of alphabetic writing systems, the typical arrangement of syllabic or abugida phoneme classes (as with Japanese kana, or devanagari), rhyme and alliteration, folk naming of correlative phoneme sets (like the "hard" and "soft" consonant groups of Slavic languages), differential learning of L2 sounds that can and cannot readily be identified with an L1 phoneme (Best, McRoberts, and Sithole, 1988), and language-determined differences in perceptual abilities (as in Trehub, 1976; Werker and Tees, 1984b; Best and Tyler, 2007) – all assure us that phonemic perception and representation are not merely a result of alphabetic writing.

The principle that adult speakers intend and perceive speech in terms of their own native set of phonemes – first observed by Kruszewski (1881) and Baudouin de Courtenay (1895), and further illustrated by Sapir (1933) and Swadesh (1934) – has been amply documented by psycholinguistic testing (e.g. Werker and Tees, 1984b; Werker and Lalonde, 1988).³ "Exemplar" and "usage-based" theories claim that phonological representations are based on multiple representations of individual acoustic forms (Pierrehumbert, 2001; Bybee, 2001), but whatever the role of exemplars, they do not make phonemic representations unnecessary (Nathan, 2007).

As Baudouin de Courtenay, its first proponent, defined it, a **phoneme** is "the psychological equivalent of a speech sound" (1895: 152). As originally conceived, the phoneme is a perceived, remembered, and intended speech sound. It is also a combination of an acoustic configuration with a set of articulations. (Later descriptions of phonemes in terms of distribution were proposed as analytic strategies, not definitions – as in Swadesh, 1934.)

A phoneme may take multiple forms in speech, but its variants or allophones are motivated, not miscellaneous. Bazell (1954) correctly observed that the reason linguists



hesitate to analyze [h] and [ŋ] as the same phoneme in English despite their complementary distribution is that there seems to be no phonetic motivation for /h/ to become [ŋ] syllable-finally, or for /ŋ/ to become [h] syllable-initially. Stampe (1987) added that there is therefore no reason for perceiving them as the same. Motivated phonological processes affect the hearer's perception of the speech of others because hearers "allow for" processes that would occur in their own speech. When an American hears another American say ['ɑɪ̃ 'wɑɪ̃ 'gou] he can interpret this as *I don't want to go* because he knows that the processes of nasalization, deletion, flapping, and vowel reduction applied by the speaker could also apply in his own speech with similar auditory results, so he can sympathetically access the speaker's intention by "undoing" these processes.

4.2 Why there is a phoneme inventory: Fortitive and lenitive processes

Studdert-Kennedy (1987) and Lindblom (1992, 2000) point out the efficiency advantages of phonemic coding, but phonemic representation is not just the result of parsimony. It results from the interaction of phonetic demands – the interaction of phonological processes of two different kinds: fortitions optimize simultaneous combinations of features (i.e. they enhance segments or "strengthen" their characteristic properties), and lenitions optimize sequences (Donegan and Stampe, 1979; Stampe, 1987).

4.2.1 Fortitions optimize segments Some processes enhance "clarity" and result in perceptually and articulatorily optimal simultaneous feature combinations (Donegan, 1978; cf. Stevens and Keyser, 1989), e.g.:

(1) DeNAS Vowels are non-nasalized.

or

(2) DeVoi Obstruents are voiceless.

These fortition processes are typically independent of segmental context,⁴ because their effect is the strengthening (maximization or optimization) of a particular phonetic property of an individual segment. DeNAS, for example, avoids the production of "extraneous" nasal formants, and thus enhances vowel quality (frontness, height, etc.); DeVoi produces a discontinuity in low-frequency energy and thus enhances an obstruent's difference from the sonorant segments that surround it. Fortitions may also reflect production advantages: non-nasal vowels maintain the speech-ready position of the velum, and voiceless obstruents obviate the articulatory adjustments needed to maintain voicing simultaneous with restricted outward airflow. The action of fortitions underlies the tendency of phoneme inventories in the world's languages to favor a particular set of segments and for children's early productions to favor those same segments.

Of course, the world's phoneme inventories are not all the same; fortitions may be limited or suppressed – i.e., the speaker may have to overcome the phonetic constraint responsible for the fortition. This acquisition of phonetic control may be complete, or it may be partial, following implicational conditions that reflect phonetic difficulty. For example, the process



- (3) De_{LAB} Vowels are non-labial

is fully realized in only a few languages – those with no labial vowel phonemes. But the full form of the process includes conditions like

- (4) $\text{De}_{\text{LAB}} \text{!LOW}$ Vowels are non-labial – especially if low.

Limited versions of this process are far more widely apparent: many languages lack low labial vowels. (Openness and roundness are phonetically incompatible.)

Fortitive processes limit speakers' perceptions of what is an intendable or "possible" speech sound. Fortitions create fewer, more inclusive categories of sounds by eliminating (potential) distinctions: for example, De_{NAS} eliminates the dual categories of nasalized and non-nasalized vowels, merging them as non-nasalized vowels. So if De_{NAS} applies in a language, it makes nasalized vowels "impossible." Similarly, $\text{De}_{\text{LAB}} \text{!LOW}$ may merge [ɒ] and [ɑ] (and [ɛ] and [æ]) as non-labial, making low labial vowels "impossible."

4.2.2 Lenitions optimize sequences Other processes result in articulatorily optimal sequences of segments, e.g.:

- (5) Na_{SIM} Sonorants adjacent to nasal consonants are nasalized

or

- (6) Vo_{SIM} Obstruents adjacent to voiced segments are voiced.

These sequence-optimizing processes affect segments in particular contexts. They are typically assimilative. They may relax the requirements for precise timing of gestures, or reduce the number or magnitude of gestures or the number or magnitude of differences between successive gestures. They may result in deletion. Because they weaken individual segments and obscure differences between adjacent segments, they are called lenitions. Note that this is not the "standard" meaning of lenition, which always involves weakening of a constriction. Some lenitive processes can actually increase a closure, as when the /z/ of *isn't*, *doesn't* assimilates to the complete closure of a following /n/, yielding, for many Americans, ['ɪdnt], ['dædnt]. Note that frequent words are most susceptible to optional lenitions. (This assimilation, for example, does not usually affect the less frequent *present*, *reason*, *cousin*, etc.)

4.2.3 Opposite motivations, opposite effects With their different motivations, fortitions and lenitions have opposite effects. Fortitions eliminate certain categories of segments, thus limiting the set of intendable or memorable sound categories. For example, in English, De_{NAS} ensures that there are no vowels which speakers mark as nasalized in lexical memory. As a result, English speakers perceive nasalized vowels as their non-nasalized counterparts, and they may produce non-nasal vowels when attempting the nasalized vowels of other languages (e.g. French *maman* [ma'mã] becomes, in English, /mama/ [ma'ma]). Alternatively, if nasality on a foreign vowel is perceived, English speakers may assume that there is a nasal consonant that accounts for it.



But of course, speakers of English do pronounce nasalized vowels, in words like *bend* and *spoon*, where the vowel is followed by a nasal consonant in the same stress group. They do so even though they typically fail to perceive any difference between the vowels of *bend* and *bed*, or those of *spoon* and *spool*. This is because the lenitive process NAS_{SIM} applies in English. Lenitive processes often create, in speech, segments that are eliminated from phonological representations – ruled out as “impossible” by applicable fortitions. The lenitive process NAS_{SIM} requires speakers to nasalize the remembered /ε/ of /bend/, so that it is actually pronounced [ẽ] in [bẽnd]. But the existence of this process also allows these speakers, as hearers, to ignore or discount this nasality; they can hear it as a result of the speaker’s submission to a process to which they are themselves subject. Thus the NAS_{SIM} process that limits their own abilities allows English hearers to assume that other speakers intend /bend/ when they say [bẽnd]. Thus, they can perceive and remember a vowel in *bend* that is non-nasal, and not ruled out by DEN_{AS} – and that is thus the same as the vowel of *bed*.

The interaction of fortitive and lenitive processes creates the phoneme inventory of a language within each individual. Both kinds of processes manifest the speaker’s inabilities. Fortitions eliminate certain sounds from the set of “intendable” sounds – the phoneme inventory. Lenitions allow the speaker to account for or ignore the actual occurrences of these “impossible” sounds (which the linguist calls allophones). The phonemes of a language are the fully specified, intendable sounds of the language, which are perceived and remembered by its speakers. They are the segments that are neither eliminated from the set of possible sounds by a fortition nor attributable to a context-sensitive lenition (Stampe, 1987; Donegan, 1995).

5. Children’s Perceptions Develop toward Adult Representations

The question of how and when a child acquires adult-like or phonemic perception occupies an important place in the literature on phonological acquisition, but it has remained a bit of a puzzle. Production abilities are not a good indicator. A child may have a large vocabulary of words and sentences she perceives and remembers, although when she begins to say them, her production abilities are limited, e.g. [pu] for *spill*, *peel*, *pail*, *pole*, *pearl*, *pull*, *pool*, *spoil* (Velten, 1943). A child can clearly perceive and remember the adult forms even of many words she doesn’t use (Velten noted that his daughter at two years could follow instructions in French and Norwegian, as well as in English, though she rarely used these languages in speaking).

5.1 Earlier and later perception

Infants appear to be born with the ability to distinguish all the vowel and consonant contrasts that appear in languages, whether or not they appear in the ambient language (Eimas, Einar, Siqueland, Jusczyk, and Vigorito, 1971 and others; Werker, 1991 reviews this research). Important changes in this ability occur between six and twelve months. Polka and Werker (1994) found that, at about this age, infants lose the ability to discriminate non-native vowel contrasts. A bit later, they lose the ability to discriminate



non-native consonant contrasts (Werker and Tees, 1984a), leading to perception largely in terms of the sounds of the ambient language.

Although it seems clear that, by about 12 months of age, children begin to respond only to differences that are distinctive in the ambient language, there is widespread reluctance to call the year-old child's perceptions phonemic. In terms of structuralist analytic methodology, complementary distribution and phonetic similarity are regarded as the crucial criteria for phonemic analysis, so investigators are often unwilling to attribute phonemic status to differences that do not distinguish minimal pairs. Werker and Pegg (1992), for example, could not establish that minimal pairs are distinguished by children under 19 months, so, instead of "phonemic perception," they refer to "language-specific phonetic" perception. Yet the similarity to adult perception is hard to ignore, and the structuralist criteria for discovering probable phonemes constitute neither a definition of the phoneme nor an explanation of the existence of such units. Structuralist characterizations of phonemes in terms of complementary distribution and minimal pairs were proposed as analytic strategies, not as definitions (Swadesh, 1934). As seen here, phonemic representations emerge from the interaction of fortitive processes that limit the universe of intendable, perceivable sounds and lenitive processes that account for deviations.

Adults seem to disregard differences that are not phonemic, rather than actually losing perceptual ability. We may assume that the child begins to do the same. But how does the year-old child know which phonetic differences may be ignored, and what phonemic (or "language-specific phonetic") contrasts are present in the adult language? For example, how does the learner of Hindi discover that he must pay attention to, and remember, stop aspiration and retroflexion, and vowel nasalization, while the learner of English either does not learn to notice (and remember) these features – or learns not to notice them?

On encountering a new language, the linguist records highly detailed phonetic forms, analyzes distribution, does away with some details, and evaluates alternative analyses. But few would claim that children do this. Yet the year-old infant perceives – and thus remembers – only some sound differences, and thus, presumably, not all the phonetic details of all the variants he hears. Which details does he represent, and how does he come to remember just these?

5.2 *Features in child representation*

In the view of features outlined above, "feature analysis" of utterances begins when the child makes a connection between an acoustic pattern and an articulatory one. For example, any articulatory gesture that includes complete oral closure (and release) results in an interval of silence or very low amplitude, followed by an abrupt onset of energy, while an incomplete oral closure (and release) results in more sound during the constriction and a less abrupt increase in amplitude at its release. The child who makes this association has discovered the feature [continuant].

In contrast, Menn and Vihman (2011) identify the acquisition of a feature with the appearance of a contrast or the reuse of a given articulation in a child's own speech. But even children who lack minimal or near-minimal pairs and who do not reuse segments or syllables across items have made relevant articulatory-auditory connections. A child may not always attempt words that require the same features, and even in multiple



attempts at the same word different features of the adult form may be reproduced. (Children have to learn that consistent substitutions help in being understood.) Menn and Vihman offer transcriptions of the first words of 50 children, and close examination reveals at least partial resemblance of each child form to the adult form. This is also true for the early forms cited by Smith (1973), Leopold (1939), and others. Children almost always reproduce some features of the adult form,⁵ and (non-)occurrence of minimal pairs may simply be a matter of chance.

Evidence that phonetic qualities of the ambient language appear even in babbling (Oller and Eilers, 1988; de Boysson-Bardies, Sagart, and Durand, 1984; Whalen, Levitt, and Wang, 1991; and others) indicates that babbling is to some degree imitative, and that the child is making connections between her own vocal noises and the production of words or phrases she hears.

Knowing the acoustic results of some articulatory gestures may allow the child to draw, from their acoustic effects, conclusions about gestures that she cannot yet perform (for example, she may know that [e] requires a tongue-fronting gesture and a non-low jaw position – like those required for [i] but less extreme – without being able to achieve this intermediate target with any reliability. And from the motor-kinesthetic-auditory linkings she knows, she may draw conclusions about feature combinations that she cannot yet produce; she may realize, for example, that [l] is sonorant, voiced, and coronal, and that it has a special auditory property (which we call [lateral]) that she cannot yet produce. This would mean that a process that eliminates laterality is overcome in perception, though it still applies to her productions.

5.3 *Phonological processes and phonemic perception*

In babbling and vocalizing, the child learns what combinations of gestures emphasize or attenuate each other's effects, and which simultaneous combinations are more easily performed. "Simultaneity constraints," which occasion fortitive processes, limit his productions to these optimal combinations, so it is hardly surprising that in canonical babbling most children produce most frequently the optimal feature combinations (segment types) that are widespread among languages, or that infants' first words contain segments that they began to control in babbling, or that children who produce relatively rare sounds in their first words are those who have produced such sounds in babbling (Locke, 1983; Vihman, 1996).

The child also learns that his production abilities are subject to "sequentiality constraints" that occasion lenitive processes, so that the actual articulations he produces in sequence (and their auditory effects) do not always match his intentions. The child can assume that similar deformations affect the intentions of others. And that means that he can discount some of the simultaneous feature combinations he hears, by attributing them to these lenitions, or perhaps to random variation. (The "perceptual magnet effect," through which sounds appear to be closer in phonetic space to prototypical sounds than to non-prototypical sounds (Kuhl, 1991), may be an effect of fortitions in perception.)

The child allows each process to apply where it can, but must suppress those that cannot apply in the ambient language. Allowing some constraints to apply and marking others for elimination is the basis for admitting sounds as possible intentions, for accounting for inadmissible but occurring sounds (allophones), and thus for



creating the limited set of images (or categories) that we call the phoneme inventory (Donegan, 1995).

Some examples can illustrate this: An English learner, who hears forms like [sām] *some* and [dɒg] *dog*, may continue to allow DeNAS to apply and assume that vowels are intended as non-nasalized. She can assume that the nasalized vowel that is heard in words like [sām] is the result of context-sensitive NAS_{SIM}, which she may also allow to continue to apply. Only if non-English forms like *[dɔ̃g] or *[sām] are heard will the child have to mark either DeNAS or NAS_{SIM} as requiring suppression. The inventory of possible sounds is thus limited to non-nasalized vowels, and vowels nasalized by adjacent consonants are perceived as their non-nasalized counterparts. (A child may of course perceive a sound quality she cannot produce. For example, a child who hears [lɛt] and [jɛs] may continue to perceive a difference in the onsets, without knowing how to produce this difference. This ability to perceive but not produce requires that she mark a neutralizing process for suppression. The child later suppresses the process in production by learning to produce [l].)

The French learner, in contrast, hears forms like [bɔ̃] *bon* ‘good, m.’ or [ʃjɛ̃] *chien* ‘dog’, and has to admit that nasalized vowels are “possible,” and that DeNAS cannot apply, since there is no nasal consonant to which the nasality can be attributed. The French learner also hears forms like [bɔn] *bonne* ‘good, f.’, or [plɛn] *pleine* ‘full’, with non-nasalized vowels before nasal consonants, and must realize that DeNAS does not apply. Further, forms like these prevent him from assuming that the nasalized vowels result from deletion of a final nasal consonant, since the final nasals remain.

Compare this to the Hindi learner, who hears forms like [hũ:] ‘am’, or [hã:] ‘yes’, and must admit that nasalized vowels are possible sounds, and that he cannot allow DeNAS to apply, because there is no adjacent nasal consonant and there is no other way to account for the nasality. When this learner hears words like [t̪ĩ:n] ‘three’ or [t̪ũm] ‘you’, he might perceive the vowel as a nasalized vowel, because it is indeed a possible sound. But in Hindi, vowels are also nasalized before nasals; in words like [kʰã:na] ‘food’, NAS_{SIM} applies. It is only when the learner makes a morphological identification of a particular nasalized vowel, for example the [ã:] of [kʰã:na], with the non-nasalized vowel of the same morpheme in another context, [kʰa: sakte] ‘can eat’ or [kʰa:ta haɪ] ‘eats’, that the nasalization is attributed to the NAS_{SIM} process and the representation of this morpheme is revised to a morphophonemic one, with a non-nasalized vowel.

So the phoneme inventory – the set of intendable feature combinations, which correspond to perceptual categories – is kept as small as possible. It is limited in two ways: (1) sounds can be ruled out by fortition processes, and (2) occurring but ruled-out sounds can be perceived as other, admissible sounds if lenition processes can account for them. Processes that can be allowed to apply *are* allowed to apply, because each process that continues to apply has a phonetic advantage for the learner. Each limitation on the inventory eases the learner’s task.

Phonemic perception is perception of sounds as intendable and producible. The neonate’s perceptions of speech sounds are not phonemic because the sounds are heard simply as sounds – not in terms of possible productions. So the infant at first hears all the distinctions of any language, e.g. s≠ʃ, b≠p, l≠j. But once the child babbles and discovers the effects of using his vocal tract, these accurate perceptions begin to be constrained by the fortitive processes that favor particular combinations. As the child begins to hear (and see) speech as producible, production constraints begin to affect his perceptions.



Fortitive processes press the advantages of limiting simultaneous feature combinations to a set of more optimal segments: e.g. $\int \rightarrow s$, $\beta \rightarrow b$, $b \rightarrow p$, $l \rightarrow j$, so that learners may assume that the more optimal segment is “intended.” Lenitive processes, which favor optimal sequences, allow learners to account for some sounds as “not intended,” so that, for example, in Japanese or Korean $[\text{ʃi}]$ may be heard as $/\text{si}/$, since $/\text{s}/ \rightarrow [\text{ʃ}]$ before $[\text{i}]$.⁶

But segments that cannot be perceived as adjusted to context continue to be heard as themselves: English *show* $[\text{ʃou}]$ will be heard by the learner as $/\text{ʃou}/$; the process that makes $[\text{ʃ}] \rightarrow [\text{s}]$ must be suppressed or at least marked for suppression, so that $/\text{ʃ}/$ is admitted as an intendable sound. Similarly, hearing $[\text{b}]$ s that are not attributable to assimilation requires the child to suppress $[\text{b}] \rightarrow [\text{p}]$ (DeVoi) in perception, even if he cannot yet produce a voiced obstruent. But $[\beta] \rightarrow [\text{b}]$ can remain active, limiting the inventory to non-implosives. The infant now reacts as if $[\text{ʃ}] \neq [\text{s}]$, and $[\text{p}] \neq [\text{b}]$, but not to $[\beta]$ versus $[\text{b}]$; $[\beta]$ is heard as equivalent to $[\text{b}]$.

Peperkamp, Pettinato, and Dupoux (2003) propose a model where complementary distribution itself affects children’s perceptions at the pre-lexical stage, but the model considers only distribution; it ignores the motivations of the processes that are responsible for the distribution. Thus it does not consider the evidence from babbling and primitive imitation that the child connects auditory stimuli with articulations and that he may consequently, like the adult, allow for the speaker’s submission to articulatory demands.

5.4 Morphophonemic representations

Phonological processes, though they often create allophones, can change the phonemic percept as well, as nasalization may change the vowel percept in Hindi. Morphophonemic representations result. These are sequences of phonemes that are in part determined by alternations. A further example would be words like German $[\text{h}^{\text{u}}\text{nt}]$ ‘dog’, which can be perceived phonemically as $/\text{h}^{\text{u}}\text{nt}/$, since no processes rule this out. And indeed *Hund* rhymes with words like *bunt* ‘colorful’. But if $[\text{h}^{\text{u}}\text{nt}]$ and $[\text{h}^{\text{u}}\text{nd}\text{əs}]$ ‘dogs’ are identified as “the same word,” the learner may realize that $[\text{h}^{\text{u}}\text{nt}]$ is really $/\text{h}^{\text{u}}\text{nd}/$. The phonetically motivated devoicing process, which applies in his own speech, would account for the $[\text{t}]$.

‘Morphophonemic’ is used here of representations like $/\text{h}^{\text{u}}\text{nd}/$, which arise through the application of phonetically motivated processes (like devoicing or nasalization). Note that such representations can arise one word at a time, without any “global” changes, general comparisons, or reference to morpheme boundaries or morpheme classes. (In contrast, “morphophonological” is here used of conventional, morphologically conditioned rules.)

6. Adults Arrive at Lexical Representations by “Undoing” Multiple Processes

Sympathetic listening applies in adult speech as well. English speakers assume, for example, that a nasalized vowel has been assimilated to a following nasal consonant,



even if no nasal consonant is evident (see, for example, Lahiri and Marslen-Wilson, 1991). Aided by phonotactic and “top-down” expectations, speakers can arrive at appropriate lexical representations even when multiple substitutions have affected the output.

The reduction of *I don't want to go* to [ˈɑɪ̃ˈwɑ̃ɪ̃ˈɡoʊ] (above) illustrates this and represents a crucial problem of speech perception – variability. Fast tempo, lack of attention or care, low prosodic saliency, situations of high redundancy, or very frequent words occasion the relaxation of process inhibitions, so that lenitive processes that ease sequences may apply more freely. Alternatively, exaggeration, emphasis or very slow tempo can occasion the application of fortitive processes that enhance particular segmental qualities. (Note that morphophonological rules are not sensitive to such prosodic or pragmatic conditions. As conventions, they apply obligatorily, exempt from the phonetic pressures of style and tempo.)

Johnson (2004) and Shockey (1974, 2003) have shown that phonologists' observations of “massive” phonological reduction are supported by instrumental analysis. Loss of whole segments or syllables can be problematic for hearers, but hearers often seem to manage, while speech recognition programs fail. Johnson reviews a number of speech recognition algorithms that attempt to cope with such reduction, but finds them inadequate. Such algorithms do not give sufficient attention to prosody, and reductions often depend crucially on prosodic factors.

It is also remarkable that recognition models do not seem to consider the allowable reduction processes of the language. Massive reduction can create homophony, but this homophony is limited by the assimilations and deletions that the language tolerates. For example, a phrase like [ˈtʰəɪ̃p ˈpʰæ̃nts] can be heard as *tight pants* or as *type pants* (as in *I don't like those tight/type pants*). But it is not heard as *tyke pants*, because English speakers do not ordinarily assimilate velars to following labials. Similarly, [ˈdʌŋ ˈkʰʌlɔ̃] can be heard as *dung color*, or as *dun color*, but not as *dumb color*, because English speakers do not ordinarily assimilate labials to following velars.

Similarly, [kʰæ̃? si] *can't see* must be perceived as /kænt si/, with a non-nasal vowel followed by a nasal consonant, because of DeNAS and NASSim and because of the (optional) process that deletes nasal consonant closure in shortening (pre-fortis) environments. And the (perceived) nasal consonant plus voiceless stop must be coronal because only a coronal stop becomes a [ʔ] (losing its oral closure) before a coronal consonant like [s].⁷

Adults arrive at speaker intentions not only by using acoustic cues (the physical realizations of phonetic representations) to determine a speaker's articulatory targets, but also by connecting the phonetic targets to the speaker's phonological intentions, via the phonological processes the language allows. Adult hearers depend on the phonological processes that are allowed in their language, with knowledge of the prosodic and pragmatic conditions under which those processes are allowed to apply, to “restore” reduced forms to a phonological shape that can fit with their top-down expectations.

7. A Note on Morphophonology

Morphophonology (or “lexical phonology”) might also be regarded as “emergent,” but it arises historically, across generations. When learners cannot account for their



elders' alternations phonetically, they may simply adopt them as conventions. This often happens when a phonetically motivated process causes an alternation, and the motivation is later obscured by an additional process. As long as the additional process applies variably, learners may "undo" it and arrive at representations like those of their elders. But if the obscuring process becomes obligatory, and the alternation becomes completely opaque to the learner, morphophonological rules arise and representations may change.

German umlaut is a simple example: vowels were fronted by a following palatal, so /mus-i/ 'mouse, pl.' was pronounced [mysi]. Later, the palatal suffix variably lost much of its palatal quality, but it remained lexically /-i/. Umlaut fronting still applied, but the suffix was simultaneously reduced, so [mysi] ~ [myse]. Learners who sometimes heard final [-i] could still represent this as /mus-i/, assuming a reduced /-i/. But when the reduction of the final vowel became obligatory and learners heard *only* [myse], they could not analyze the [y] as a fronted /u/, attributable to a following /i/. To admit /y/, they had to assume that the invariant [ɛ] suffix was /ɛ/ and that the [y] was intended, i.e. that it was phonemic. They had to suppress the processes that delabialized front vowels and de-palatalized labial vowels. And they had to create a rule to front the /u/ to /y/ in words like /mus/ when the plural suffix /ɛ/ was added. Rules of this sort are not phonetically motivated, but entirely conventional. They may generalize on morphological grounds, as umlaut did, but they do not represent constraints on speaker abilities.

8. Conclusion

In vocalization and babbling, the learner develops knowledge of his own articulatory abilities and their limitations. Experience brings the implicit knowledge that some combinations of features (simultaneous or sequential) present difficulties, and that these can (or must) be avoided by substituting forms that lack the difficulty. The application of fortitive processes keeps the inventory of phonemes relatively small, compared with the range of sounds humans can make. The learner maintains all possible fortitions, because perceiving a sound as an intention of the speaker would be admitting that one has to learn to produce it on purpose.

But since there are also limitations on possible sequences of sounds, and since lenitive substitutions result, sounds that are not members of the basic inventory occur as variants, or allophones. Learners can disregard allophonic alternation or variation – not because they have analyzed the distribution of allophones, but because they experience the phonetic motivations for the allophonic differences.

The "learnability problem" in phonology arises from the learner's apparent need to posit both forms and processes, and from the fact that the forms and processes depend on each other. But if processes are discovered as part of learning to use one's physical capabilities, the circularity is broken. The learner's task is to determine, on the basis of the spoken forms, which processes must be inhibited.

In the model of phonology presented here, perception and production are closely related. Both refer to the same set of processes. There is a production bias in perception, since perception is constrained by production processes, both fortitive and lenitive.



Learners want to perceive in terms of articulatorily optimal categories, and they recognize articulatory motivations for divergence from these categories. Production, in turn, has a perception bias: fortitive processes enhance perceptual properties of individual segments, and the limitations on lenitive processes are also perceptually motivated.

Phonology, as the system that governs the perception, representation, and production of speech, is inseparable from phonetics, and it emerges from speaker abilities (and inabilities). It might be argued, of course, that the conventional patterns of morphophonology are also emergent, since these conventional patterns may arise from natural substitutions whose motivations become obscured by additional motivated changes. But this emergence is historical. Morphophonology is learned purely by observation of alternations in the speech of others, and changes of lexical forms occur during speaker-to-speaker transmission (Donegan, 1993; Donegan and Nathan, *in press*). The alternations that constitute the living phonology of a language, in contrast, emerge from the learner's/speaker's/hearer's own limitations, both perceptual and productive. Phonological substitutions are responses to these limitations, and they do not merely make speech grammatical; they make it pronounceable, perceivable, and memorable.

By focusing on phonology as a set of automatic responses to phonetically motivated constraints, we may see a path from early language-general perceptual accuracy to language-specific perception, as well as a path from babbling to adult speech. We may also better understand the ways in which adult speakers are able to identify lexical items in the continuous and continuously varying speech signal.

NOTES

- 1 For further discussion and examples of differences between lexical or morphophonological rules and automatic or "natural" processes, see Donegan and Stampe, 1979, 2009.
- 2 At the time, interest in formalism and the Chomskyan view of "innateness" prevented generative phonologists from seeing phonology as arising from a system based in human physiology and perception.
- 3 In some circumstances, speakers can distinguish allophones (Best, McRoberts, and Sithole, 1988; Best and Tyler, 2007), especially if they appear "out of context" and if they differ by a feature that is distinctive elsewhere in the language (cf. Peperkamp, Pettinato, and Dupoux, 2003).
- 4 Fortitions may be "context-free," or they may be limited to specifiable prosodic (rhythmic and accental) conditions, e.g. stressed syllables, onsets (see Donegan and Stampe, 1978).
- 5 Admittedly, children occasionally make up words that resemble no adult form but have obvious meaning; for example, my daughter had a special blanket that she called [*'m̥i, m̥i*] for several years.
- 6 Simply stated, these fortitions are: sibilants are anterior, stops are non-implosive, obstruents are voiceless, coronals are non-lateral. Lenitions may include: sibilants are palatal before palatals, stops are voiced between voiced sounds, etc.
- 7 In some derivational models, these processes apply in a particular, "counter-bleeding" order (e.g. nasalization must precede nasal stop deletion). But they can equally apply *simultaneously*. Since /kænt si/ meets all the requirements of aspiration, nasalization, nasal stop deletion, and glottalization, all can apply at once. Glottalization "feeds" freely (re)applying coronal closure loss (cf. Donegan and Stampe, 1979).

REFERENCES

- Baudouin de Courtenay, Jan. 1895. *Versuch einer Theorie phonetischer Alternationen*. Strasbourg: K. J. Trübner. Translated by Edward Stankiewicz as An attempt at a theory of phonetic alternations. In *A Baudouin de Courtenay Anthology*, pp. 144–212. Bloomington: Indiana University Press, 1972.
- Bazell, Charles E. 1954. The choice of criteria in structural linguistics. *Word* 10: 6–15.
- Best, Catherine T., Gerald W. McRoberts, and Nomathemba M. Sithole. 1988. The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–60.
- Best, Catherine T. and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language Experience in Second Language Speech Learning, In Honor of James Emil Flege*, pp. 13–34. Amsterdam: John Benjamins.
- Bybee, Joan L. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- de Boysson-Bardies, Bénédicte, Laurent Sagart, and Catherine Durand. 1984. Discernible differences in the babbling of infants according to target language. *Journal of Child Language* 11: 1–15.
- Donegan, Patricia. 1978. On the natural phonology of vowels. Doctoral dissertation, Ohio State University (Ohio State University Working Papers in Linguistics 23). Published 1985, New York: Garland.
- Donegan, Patricia. 1993. On the phonetic basis of phonological change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*, pp. 98–130. London: Longman.
- Donegan, Patricia. 1995. The innateness of phonemic perception. In Vida Samiian and Jeanette Schaeffer (eds.), *Proceedings of the Twenty-Fourth Western Conference on Linguistics*, vol. 7, pp. 59–69. Fresno, CA: California State University, Department of Linguistics.
- Donegan, Patricia and Geoffrey Nathan. In press. Natural Phonology and sound change. In Patrick Honeybone and Joseph Salmons (eds.), *The Handbook of Historical Phonology*. Oxford: Oxford University Press.
- Donegan, Patricia and David Stampe. 1978. The syllable in phonological and prosodic structure. In Alan Bell and Joan Bybee Hooper (eds.), *Syllables and Segments*, pp. 25–34. Amsterdam: North-Holland.
- Donegan, Patricia and David Stampe. 1979. The study of natural phonology. In Daniel A. Dinnsen (ed.), *Current Approaches to Phonological Theory*, pp. 126–173. Bloomington: Indiana University Press.
- Donegan, Patricia and David Stampe. 2009. Hypotheses of natural phonology. *Poznan Studies in Contemporary Linguistics* 45: 1–31.
- Eimas, Peter D., Einar R. Siqueland, Peter W. Jusczyk, and James Vigorito. 1971. Speech perception in infants. *Science* 171: 303–306.
- Fry, D. B. 1966. The development of the phonological system in the normal and deaf child. In F. Smith and G. A. Miller (eds.), *The Genesis of Language*, pp. 187–206. Cambridge, MA: MIT Press.
- Johnsen, Sverre. 2012. A diachronic account of phonological unnaturalness. *Phonology* 29: 505–531.
- Johnson, Keith. 2004. Massive reduction in conversational American English. In Kiyoko Yoneyama and Kikuo Maekawa (eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the First Session of the 10th International Symposium*, pp. 29–54. Tokyo: National International Institute for Japanese Language.
- Jun, Sun-Ah. 1996. Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology* 13: 377–407.
- Kingston, John and Randy Diehl. 1994. Phonetic knowledge. *Language* 70: 419–454.
- Kruszewski, Mikołaj. 1881. *Über die Lautabweschlung*. Kazan: Universitätsbuchdruckerei. Translated by Robert Austerlitz as On sound alternation, in Mikołaj Kruskowski, 1995, *Writings in General Linguistics*, ed. Konrad Koerner, pp. 5–34. Amsterdam: John Benjamins.

- Kuhl, Patricia K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93–107.
- Lahiri, Aditi and William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38: 245–294.
- Leopold, Werner F. 1939. *Speech Development of a Bilingual Child. Volume 1: Vocabulary Growth in the First Two Years*. Evanston, IL: Northwestern University Press.
- Lindblom, Björn. 1992. Phonological units as adaptive emergents of lexical development. In Charles A. Ferguson, Lisa Menn, and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, pp. 131–163. Timonium, MD: York Press.
- Lindblom, Björn. 2000. Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. In Klaus J. Kohler, Randy K. Diehl, Olle Engstrand, and John Kingston (eds.), *Studies in Speech Communication and Language Development, Dedicated to Björn Lindblom on his 65th Birthday*. *Phonetica* 57: 297–314.
- Locke, John L. 1983. *Phonological Acquisition and Change*. New York: Academic Press.
- Locke, John L. and Dawn M. Pearson. 1992. Vocal learning and the emergence of phonological capacity: A neurobiological approach. In Charles A. Ferguson, Lisa Menn, and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, pp. 91–129. Timonium, MD: York Press.
- MacNeilage, Peter. 2008. *The Origin of Speech*. Oxford: Oxford University Press.
- Menn, Lise and Marilyn Vihman. 2011. Features in child phonology: Inherent, emergent, or artifacts of analysis? In G. Nick Clements and Rachid Ridouane (eds.), *Where Do Features Come From?*, pp. 259–302. Amsterdam: John Benjamins.
- Moosmüller, Sylvia. 2007. Phonetics needs phonology. In Viktor A. Vinogradov (ed.), *Lingvističeskaja polifonija. Sbornik v čest' jubileja professora R. K. Potapovoj*, pp. 488–503. Moscow: Jazyki Slavjanskich Kul'tur.
- Nathan, Geoffrey. 2007. Phonology. In Dirk Geeraerts and Hubert Cuykens (eds.), *The Oxford Handbook of Cognitive Linguistics*, pp. 611–631. Oxford: Oxford University Press.
- Oller, D. Kimbrough and Rebecca Eilers. 1988. The role of audition in infant babbling. *Child Development* 59: 441–449.
- Peperkamp, Sharon, Michele Pettinato, and Emmanuel Dupoux. 2003. Allophonic variation and the acquisition of phoneme categories. In Barbara Beachley, Amanda Brown, and Frances Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*, vol. 2, pp. 650–661. Somerville, MA: Cascadilla Press.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Lexical Structure*, pp. 137–157. Amsterdam: John Benjamins.
- Polka, Linda and Janet F. Werker. 1994. Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–435.
- Sapir, Edward. 1933. La réalité psychologique des phonèmes. *Journal de psychologie normale et pathologique* 30: 247–265. Reprinted 1949 from the author's MS as The psychological reality of phonemes, in *Selected Writings in Language, Culture, and Personality*, ed. David Mandelbaum, pp. 46–60. Berkeley: University of California Press.
- Shockey, Linda. 1974. Phonetic and phonological properties of connected speech. Doctoral dissertation, Ohio State University (Ohio State Working Papers in Linguistics 17, pp. 1–143).
- Shockey, Linda. 2003. *Sound Patterns of Spoken English*. Oxford: Blackwell.
- Smith, Neilson V. 1973. *The Acquisition of Phonology: A Case Study*. Cambridge: Cambridge University Press.
- Stampe, David. 1969. The acquisition of phonetic representation. In Robert I. Binnick, Alice Davison, Georgia Green, and Jerry L. Morgan (eds.), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, pp. 443–454. Chicago: Chicago Linguistic Society.
- Stampe, David. 1973. A dissertation on natural phonology. Doctoral dissertation, University of Chicago. Published 1980 with Afterthoughts. New York: Garland, and Bloomington: Indiana University Linguistics Club.

- Stampe, David. 1987. On phonological representations. In Wolfgang U. Dressler, H. C. Luschutzky, O. E. Pfeiffer, and J. R. Rennison (eds.), *Phonologica 1984*, pp. 287–300. Cambridge: Cambridge University Press.
- Stevens, Kenneth and Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65: 81–106.
- Studdert-Kennedy, Michael. 1987. The phoneme as a perceptuomotor structure. In Alan Allport, Donald G. MacKay, Wolfgang Prinz, and Eckart Scheerer (eds.), *Language Perception and Production*, pp. 67–84. New York: Academic Press.
- Swadesh, Morris. 1934. The phonemic principle. *Language* 10: 117–129. Reprinted 1957 in Martin Joos (ed.), *Readings in Linguistics*, pp. 32–37. Washington, DC: American Council of Learned Societies.
- Trehub, Sandra E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47: 466–472.
- Velten, H. V. 1943. The growth of phonemic and lexical patterns in infant language. *Language* 19: 281–292.
- Vihman, Marilyn M. 1996. *Phonological Development*. Oxford: Blackwell.
- Werker, Janet F. 1991. The ontogeny of speech perception. In Ignatius Mattingly and Michael Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*, pp. 91–116. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Werker, Janet F. and C. E. Lalonde. 1988. Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology* 24(5): 672–683.
- Werker, Janet F. and Judith E. Pegg. 1992. Infant speech perception and phonological acquisition. In Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, pp. 285–311. Timonium, MD: York Press.
- Werker, Janet F. and Richard C. Tees. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development* 7: 49–63.
- Werker, Janet F. and Richard C. Tees. 1984b. Phonetic and phonemic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75: 1866–1878.
- Whalen, Douglas H. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18: 3–35.
- Whalen, Douglas H., Andrea G. Levitt, and Emily Q. Wang. 1991. Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, 18: 501–516.