Published in final edited form as:
J Phon. 2010 January 1; 38(1): 109–126. doi:10.1016/j.wocn.2009.09.001.

Perception of initial obstruent voicing is influenced by gestural organization

Catherine T. Best[1] and
MARCS Auditory Laboratories, University of Western Sydney, Penrith NSW 1797, Australia and Haskins Laboratories, 300 George St., New Haven CT 06511, U.S.A.

Pierre A. Hallé[2]
Laboratoire de Phonétique et Phonologie, 19 rue des Bernardins, 75005 Paris, France, Laboratoire de Psychologie et Neurosciences Cognitives, CNRS-Paris 5 and Haskins Laboratories, 300 George St., New Haven CT 06511, U.S.A.

Abstract

Cross-language differences in phonetic settings for phonological contrasts of stop voicing have posed a challenge for attempts to relate specific phonological features to specific phonetic details. We probe the phonetic-phonological relationship for voicing contrasts more broadly, analyzing in particular their relevance to nonnative speech perception, from two theoretical perspectives: feature geometry and articulatory phonology. Because these perspectives differ in assumptions about temporal/phasing relationships among features/gestures within syllable onsets, we undertook a cross-language investigation on perception of obstruent (stop, fricative) voicing contrasts in three nonnative onsets that use a common set of features/gestures but with differing time-coupling. Listeners of English and French, which differ in their phonetic settings for word-initial stop voicing distinctions, were tested on perception of three onset types, all nonnative to both English and French, that differ in how initial obstruent voicing is coordinated with a lateral feature/gesture and additional obstruent features/gestures. The targets, listed from least complex to most complex onsets, were: a lateral fricative voicing distinction (Zulu /ɬ/-/ɮ/), a laterally-released affricate voicing distinction (Tlingit /tɬ/-/dɮ/), and a coronal stop voicing distinction in stop+/l/ clusters (Hebrew /tl/-/dl/). English and French listeners' performance reflected the differences in their native languages' stop voicing distinctions, compatible with prior perceptual studies on singleton consonant onsets. However, both groups' abilities to perceive voicing as a separable parameter also varied systematically with the structure of the target onsets, supporting the notion that the gestural organization of syllable onsets systematically affects perception of initial voicing distinctions.

Index Terms

nonnative speech perception; voicing contrasts; phonological features; articulatory gestures; syllable onsets

1. Introduction

In this paper, an experimental cross-linguistic approach is employed to compare how featural and gestural accounts of phonological structure could handle key aspects of language-specific speech perception. We focus on perception of voicing in syllable onsets of varying segmental

[1]c.best@uws.edu.au
[2]pierre.halle@univ-paris5.fr

and/or articulatory complexity for two reasons. First, laryngeal distinctive features, especially relating to contrasts in consonant voicing, are found in phonological descriptions of virtually all spoken languages. Any phonological theory thus has to deal successfully with voicing distinctions. Second, despite the wide impact of Abramson and Lisker's (1970) seminal study on nonnative stop voicing perception, reports on the perception of voicing in nonnative contrasts remain fairly small in number (e.g., Williams, 1977). While that core set of studies supports Abramson and Lisker's classic findings (for recent work, see Beach, Burnham & Kitamura, 2001; Sharma, Marsh & Dorman, 2000; Shin, 2001), its nearly-exclusive restriction to initial singleton plosive stops offers limited perspective on phonetic-phonological relationships in perception. The present investigation of voicing distinctions in onsets of varying complexity can thus advance understanding of nonnative voicing perception, as well as addressing the broader theoretical issue of whether and how current phonological approaches can account for voicing perception. We consider one representative each for the featural and the gestural approaches to phonology, respectively, feature geometry (Clements, 2003, 2005, 2006; Clements & Hume, 1995) and articulatory phonology (Browman & Goldstein, 1986, 1989, 1992).

1.1. Implications of featural and gestural phonology for speech perception

The relative merits of the two phonological approaches have primarily been evaluated with respect to output phonology (Clements, 1992, 2005): Does one yield a more principled and/or more economical account of the various phonological systems of the world's languages? Which provides a more straightforward account of the phonological processes governing, for example, context-dependent reorganization? The gestural and featural approaches also have been compared on how phonological representations interface with surface phonetic forms. Whereas the feature approach implies a computational interface between phonological representation and phonetic realization, the need for such an interface is circumvented in articulatory phonology, in which the gestural representation *is* the direct specification of the produced gestures. Such questions have been the subject of long-term debate, with many more contributions from the featural perspective (e.g., Clements, 2003; Ewen & van der Hulst, 2001; Keating, 1988, 1990; Vaux, 1998), making it fairly familiar as compared to the gestural perspective (Browman & Goldstein, 1992; Goldstein, Byrd, & Saltzman, 2006; Goldstein & Fowler, 2003). In this paper, however, we do not compare the two approaches with respect to output phonology or putative interface with phonetics. Instead, we focus on the implications for speech perception of the representational structures proposed, that is, feature versus gesture structures.

One crucial premise of the gestural approach, directly relevant to our focus on perception, is that gestures are the primitives of speech perception as well as of speech production (e.g., Best, 1984, 1994; Fowler, 1986, 1989; Goldstein & Fowler, 2003). Consistent with that view, recent research provides new evidence of perceptuo-motor compatibility between speech perception and production (Fowler, Brown, Sabadini, & Weihing, 2003; Galantucci, Fowler, & Goldstein, 2009; but see Mitterer & Ernestus, 2008). Note that positing speech gestures as workable primitives for speech perception, production, and phonological description is more radical, and more parsimonious, than the rather underspecified claim that speech perception and production share common mental representations and processes.

The non-conventional view of articulatory gestures as the common currency of both perception and production is the position taken in PAM, the model of nonnative speech perception developed by Best and colleagues (Best, 1995; Best & Tyler, 2007), which motivates the present perceptual examination. PAM explicitly assumes that gestural rather than featural properties inform speech perception, and that, at the segmental level, nonnative speech perception is driven by both the phonological organization of the listener's native language,

and by perceived articulatory similarities between nonnative phones and native phonemes. Data accumulated thus far in cross-language experiments of speech perception are largely consistent with PAM's premises with respect to phonological structure and perceptual similarities (e.g., Best, McRoberts, & Goodell, 2001). While it is possible that they might also be consistent with a featural approach to speech perception, the implications of the featural approach for speech perception have not been tested, or even clearly formulated. The present study is designed to specifically investigate this as yet under-examined issue.

1.2. Features vs. gestures: Timing representation, contextual dependencies

The featural and gestural descriptions of speech do share certain common assumptions, for instance, that spoken utterances display hierarchical organization (see Clements, 1992; also see Clements, 2003, 2005). They differ, however, in critical aspects, such as the fact that gestures are dynamic and have temporal and spatial extent, whereas features are static and lack internal structure (Browman & Goldstein, 1989; Clements, 2003, 2006; Goldstein, Byrd, & Saltzman, 2006). In this paper, we focus on two crucial theoretical differences between the phonological primitives assumed, respectively, by featural and gestural accounts. One crucial difference concerns the temporal dimension. In featural approaches, feature structures are attached to segments, with possible specifications of association links indicating "simultaneous features." No further detail on temporal organization is specified. Contour stops may be viewed as an exception to this lack of specification. They can be described, for example, as a sequence of aperture nodes -- in the case of affricates, a stop closure node and a fricative release node (Steriade, 1993). But the need for such "contour specification" is controversial among feature phonologists. For example, a feature geometry proposal is that both simple and affricate stops are non-contoured stops, that is, have no internal sequencing, with [+strident] specifying the release of affricates (Clements, 1999, 2006; Kehrein, 2002).

In stark contrast to features, gestures have intrinsic timing and specifically involve phasing relationships to one another. Indeed, intergestural phasing is central to articulatory phonology. In recent developments of articulatory phonology, intergestural phasing is posited to be controlled by a system of coupled oscillators that can settle into two intrinsically stable modes, in which the onsets of two (or more) gestures are either *in-phase* or *anti-phase* (Goldstein, Chitoran, & Selkirk, 2007). Both in-phase and anti-phase couplings are more stable than intermediate or indeterminate phasing relationships. The in-phase mode is more stable than the anti-phase mode, and is sufficient to provide a temporal structure for CV syllables, in which the C and V gesture onsets are in-phase, that is, launched at the same time. This conceptualization readily accounts for the cross-linguistic prevalence of CV syllables with singleton consonant onsets, relative to the lower incidence of CVC syllables, in which the onset and coda Cs are anti-phased, as well as to the lower incidence of syllables with cluster onsets, which involve anti-phased gestures that compete for an in-phase relationship with the V nucleus. In other words, simpler timing structures are preferred over complex ones.

The intergestural phasing approach implies accurate alignment of multiple gestures in a CV onset. In particular, gestural phasing within syllable-initial Cs is crucial for laryngeal contrasts that have been characterized in terms of voice onset time (VOT) distinctions: the "basic" three-way distinction among prevoiced, voiceless unaspirated, and voiceless aspirated stops (e.g., [d]-[t]-[t^h]). There is good agreement between articulatory phonology and VOT descriptions, in that VOT parallels intergestural timing specification. On the featural side, descriptions of laryngeal contrasts are not related to the VOT parameter, which is viewed as an acoustic characteristic whose value results from the phonetic implementation of the feature representation, rather than as a defining parameter in itself. Laryngeal contrasts are not easily amenable to feature descriptions. This is particularly true for English obstruents, for which the use of [spread glottis] as a defining feature versus an enhancement feature is still debated

(Clements, 2003; Halle & Stevens, 1971; Keating, 1990; Keyser & Stevens, 2001; Vaux, 1998), and the feature description of the voicing contrast is unavoidably heterogeneous, from a phonetic viewpoint, across stops and fricatives. More generally, the substantial variations in phonetic detail of voicing contrasts across languages, segment types, and syllable positions has led some phonologists to abandon attempts to correlate voicing feature distinctions with unique phonetic properties (e.g., Keating, 1984). To summarize, features and gestures differ wildly with respect to their treatment of timing. The case of voicing contrasts[3] provides a neat illustration of this difference.

The two approaches also differ with respect to the dimensionality of their representations, although this point is more nuanced. Featural systems try to eschew "dependencies" among features, which thus behave like independent atoms in feature descriptions (see, for example, how proposed feature geometries evolved from the Halle-Sagey model [Halle, 1995; Sagey, 1986] to Halle, Vaux and Wolfe [2000], the latter proposal dispensing with the dependencies of [back, high, low] on [dorsal]; cf. Lahiri and Reetz, this volume, for a brief review of feature geometries). The relative independence of individual features is perhaps most easily understood in asynchronous feature matrix descriptions (Stevens, 1996, 2002), where features are not necessarily attached to segments, and individuating the value of a given feature, for example [voice], is unconstrained by the values of other features. In hierarchical descriptions such as the widely accepted feature geometry proposed by Clements and Hume (1995), features are linked to one feature-class node (or possibly two when describing assimilation processes), and can be associated with a segment's root node or with a skeletal position in the prosodic tier, but these links do not constrain a given feature value.

It is not clear, however, whether and how the feature structure of underlying representations would be expected to impact on speech perception. Although perception is not often discussed in phonological research (but see relevant work by Lahiri and colleagues: Lahiri & Reetz, this issue), commonly encountered statements such as "feature F is extracted from the surface by the perceptual system" suggest that, in the featural approach, the surface level of representation in the output can be taken as the input to perception. Thus, this level of representation should logically maintain a relative independence of any feature value from other feature values. It is in that sense that feature units may be viewed as independent.

In contrast, gestures "…are 'larger' units than features – consisting of bundles of information … that in feature theory are analyzed as independent features" (Clements, 1992). The "gestural scores" used in articulatory phonology consist of *interdependent* gesture units that are linked by phasing relations, as we have noted. For features, specified associations between nodes at the feature, feature-class, or root level, are much less constraining than are the phasing relations for gestures. What may be the perceptual implications of independent feature units versus highly interdependent gesture units? A recent extension of articulatory phonology directly addresses a related issue of interdependent articulatory gestures, or contextual dependency: the articulatory organ hypothesis (AOH) proposes that "within-organ" distinctions produced by differentiated actions of a single articulator pose greater perceptual difficulties than "between-organ" distinctions.[4] Thus, on the AOH view, contextual dependency increases perceptual difficulty. Extending the logic of contextual dependency, the gestural view should predict

[3]We use "voicing" as a cover term for feature distinctions covered by [voice] and [spread glottis] or, on the gestural side, for specifications of the phasing between a glottal opening-closing gesture and supralaryngeal constriction gesture. We do not include [constricted glottis], or glottal gestures other than opening-closing, which mainly refer to laryngeal adjustments related to implosives and ejectives.

[4]The AOH was developed to account for infants' attunement to the native language, predicting an earlier and/or more precipitous developmental decline, and/or a need for experience-based induction in discrimination of within-organ than between-organ contrasts (Studdert-Kennedy & Goldstein, 2003). English-learning 11-month-olds did indeed show a decline, relative to 7-month-olds, in discriminating Zulu [ɮ]-[ɬ], [k]-[k′], [b]-[ɓ] *and* native English /s/-/z/, which all involve within-organ laryngeal distinctions; they showed no decline for Tigrinya between-organ [p′]-[t′] (Best & McRoberts, 2003).

increasing perceptual difficulty with increasing gestural complexity. In contradistinction, on the featural view, independence of feature values predicts a more uniform analysis, irrespective of the target's structural complexity and essentially insensitive to internal timing.

1.3. Perception of voicing in nonnative onsets: The possible role of onset complexity

We have argued that voicing offers a promising focus for perceptual comparisons of the featural and gestural approaches especially because voicing seems directly linked to timing in the speech domain. Voicing is also interesting to investigate because the supralaryngeal gestural context in which the relevant glottal gestures are achieved can be easily manipulated in contrasting languages that differ with respect to that context, as we will describe shortly. From a feature point of view, the value of, for instance, [voice] should not be blurred by articulatory complexity, given the assumed independence of feature values. By comparison, from a gesture standpoint, swift detection of the relevant laryngeal gesture, glottal opening-closing,[5] could well be influenced by the co-produced gestures at other, supralaryngeal articulatory tiers, given the assumed interdependence of gestures.

These predictions are consistent with several reports on discrimination and categorization of voicing contrasts. For example, Werker and colleagues (Tees & Werker, 1984; Werker, Gilbert, Humphrey & Tees, 1981) found that English listeners discriminated Hindi breathy-voiced versus voiceless aspirated dental stops /d^{h}/-/t^{h}/[6] rather poorly.[7] Featurally, [spread glottis] is specified for both onsets, which contrast in [voice]. A possible featural explanation of the poor discrimination performance, then, might be that [voice] is nonnative to English phonology, where only [Ø]-[spread glottis] is used to specify voicing, for which these two phones have the same setting. An alternative possibility is that each has two laryngeal feature specifications, [±voice] plus [spread glottis], a nonnative combination for English. Neither possibility, however, offers an entirely satisfactory account of the perceptual finding. According to the gestural approach, by comparison, Hindi /d^{h}/ and /t^{h}/ both involve a glottal opening-closing gesture, which differs in phasing relative to the tongue tip constriction gesture (respectively, in phase with the closure peak of [d̪^{ɦ}] vs. in phase with the closure release of [t̪^{h}]). Thus, it is a nonnative within-organ contrast, predicted by the AOH to be difficult.

Another likely case of gestural context effects in perception can be found in a recent study on, among other contrasts, the Zulu lateral fricative voicing distinction /ɮ/-/ɬ/, which involves greater gestural complexity than, for example, the simple coronal fricatives /z/-/s/. The supralaryngeal gestures for lateral fricatives comprise a coronal constriction (tongue tip) and a dorsal constriction that permits lateral airflow (tongue body). English listeners showed very good discrimination of the Zulu voiced-voiceless lateral fricatives (Best et al., 2001). However,

[5]Goldstein and Browman (1986; Browman & Goldstein, 1986) proposed that voicing distinctions are signaled by presence/absence of a glottal opening-closing gesture that occurs in some specified phase relation with a supralaryngeal (oral) constriction gesture. This offers a unified physical basis for voicing contrasts across and within languages. In particular, the English phonologically voiced (e.g., /b/) but phonetically voiceless unaspirated (i.e., [p]) initial stops can be described as lacking a glottal opening-closing gesture and realizing a critical degree of glottal adduction *at release* of oral closure. This description allows direct comparison not only to glottal gestures for English voiced fricatives, but also to the phonologically and phonetically fully voiced stops of French (the “pre-voiced” stops), which also lack a glottal opening-closing gesture but maintain a critical degree of glottal adduction throughout *both* closure and release of the oral constriction. In contrast, French voiceless stops, in spite of their *acoustic* similarity to the VOT of English phonologically-voiced initial stops, are produced with a glottal opening-closing gesture that is in phase with oral release, and thus are gesturally different from the English phonologically-voiced stops of roughly the same VOT value.

[6]Depicting segments as abstract phonological entities (i.e., as /x/) versus as phonetically detailed entities (i.e., as [x]) is admittedly a vexed issue in the literature, where there is both disagreement and inconsistency of usage. For our purposes in this paper, we will use a “phonological” designation (/x/) to refer to abstract entities and/or contrastive relationships, but will use a narrower “phonetic” designation ([x]) when presenting or hypothesizing more detailed phonetic (articulatory/acoustic) characteristics. It is important to certain arguments of the present paper that we signify the two levels of description differently.

[7]It is not known how these phones were categorized, because only discrimination was assessed. However, articulatory and acoustic similarities and dissimilarities to English aspirated alveolar /t/ suggest that English listeners might have heard both Hindi dentals as deviant versions of native /t/.

while the listeners could have straightforwardly perceived them as a simple English fricative voicing contrast (e.g., /ð/-/θ/), they instead gave highly varied open-response categorizations to English consonant onsets, which included /l/ for [ɮ] targets but were often phonotactically impermissible sequences of fricatives, stops or affricates plus /l/, rather than single-segment fricatives; some even lacked a voicing contrast (e.g., /ʒ/ vs. /ʒl/). There was one constant, though: everyone heard *some* phonological distinction, with at least one item being heard as having a cluster onset. In other words, English listeners perceived articulatorily complex onsets, and they did not always extract the "correct" voicing of the onsets they were presented with, suggesting that onset complexity can pose a challenge to the perception of voicing. As we have argued, context-dependent effects in perception are conceivably more compatible with gestural than featural representations. (In the case of /ɮ/-/ɬ/, extraction of the [voice] feature value should not depend on the presence/absence of [lateral].)

A further issue is whether speech perception may be affected simply by segmental or supralaryngeal complexity as specified in terms of number of involved gestures. Recent data on the perception of Hebrew stop + liquid clusters (/tl, dl, kl, gl/, etc.) by English and French listeners showed that, despite the fact that these onsets are segmentally complex, and that the place of articulation of the coronal stops preceding /l/ was systematically misperceived, voicing was identified accurately (Hallé & Best, 2007). This suggests that factors other than segmental complexity affect the perception of voicing. One such additional factor could be the specification of the coupling, or phasing, between the laryngeal gesture and the supralaryngeal gestures.

1.4. The present study

The studies reviewed above were not designed specifically to test the possible perceptual impact of onset complexity in terms of overall featural or gestural organization. The present study is therefore designed to directly address this issue by examining the perception of three types of onsets that vary in structural complexity. Two of these were the lateral fricatives of Zulu and dental-lateral clusters of Hebrew already discussed. The third was the lateral affricates /dɮ/-/tɬ/ of Tlingit, an indigenous Alaskan language (Athabaskan family), for which the voicing contrast is one of [Ø]-[spread glottis], that is, an aspiration distinction (Ladefoged & Maddieson, 1996; Maddieson, Smith & Bessell, 2001). All three target onset types combine a voicing contrast (laryngeal) with a coronal constriction and a lateral gesture (supralaryngeal).[8] Thus they share a common subset of gestures, but differ in how they organize those gestures.

Schematic gestural scores for these three onset contrasts are shown in Figure 1. The Hebrew cluster onsets (Figure 1A) are segmentally complex in that they involve two consonants in sequence, phonologically represented as /dl/-/tl/. Therefore, the in-phase relationship between the oral gestures and the critical glottal gesture for the voiceless cognate is assumed to be restricted to the initial stop. Moreover, the stop and the lateral are anti-phased to one another to maintain their sequentiality; they are thus in competition to be in-phase with the vowel nucleus. This shifts the onset of the stop gesture leftward and that for the lateral rightward, such that the vowel is in-phase with the center of the onset cluster (C-center: see Goldstein et al., 2006;Nam et al., 2009). The supralaryngeal articulation of the singleton Zulu lateral fricatives (Figure 1B) is posited to consist of two in-phase components, a midline tongue tip closure (coronal), and a critical tongue body constriction that allows turbulent airflow around the sides of the tongue (lateral). For voiceless /ɬ/, the peak of a glottal opening-closing gesture is phase with the peak of the fricative constriction (peak-phasing). By contrast, critical glottal adduction, resulting in glottal pulsing, is maintained throughout voiced /ɮ/, that is, there is no

[8]Interestingly, orthographic representations (Zulu, Tlingit) and Roman transliterations (Hebrew) of these onset contrasts are almost identical: 'dl' versus 'tl', except for the Zulu voiceless fricative /ɬ/, which is 'hl' in written Zulu.

glottal opening-closing gesture. Thus, the gestural structure is clearly less complex, and the overall intergestural phasing is tighter, for the Zulu fricatives /ɮ/-/ɬ/ than for Hebrew clusters /dl/-/tl/. The Tlingit lateral affricates /dɮ/-/tɬ/ (phonetically, [t͡ɬ]-[t͡ɬʰ]), in turn, offer even tighter intergestural phasing than the Zulu lateral fricatives (see Figure 1C). They are single segments like the fricatives, and their gestural components are similar to those of the Hebrew clusters. They employ a coronal stop closure and a critical constriction degree for the lateral release of the closure. Yet their intergestural structure differs from both the fricatives and the clusters: the glottal gesture for the voiceless affricate is assumed to be in-phase with both the stop and lateral gestures. The Tlingit contrast also differs from the Hebrew and Zulu targets in that the glottal opening-closing gesture for the voiceless item is in-phase with closure release (as in English stop voicing distinctions).[9] Because all gestures are in-phase at the onset, unlike the peak-phased glottal gesture for the voiceless Zulu lateral fricative, the Tlingit lateral affricates show the tightest intergestural coupling of the three target contrasts. However, they are intermediate in supralaryngeal gestural complexity between the Zulu fricatives and the Hebrew clusters. Table 1 summarizes how the contrasts vary in supralaryngeal complexity and intergestural phasing of supralaryngeal and glottal gestures.

Figure 2 presents the posited feature trees for the three onset voicing contrasts. We have included a single feature tree for each language, because the only setting that differs between the paired onsets for each language is voicing. Thus the feature settings for each target language for the phonological voiced/voiceless distinction are designated as either [+voice] vs. [-voice], or Ø vs. [spread glottis], consistent with our preceding discussions (these distinctions are boldfaced in the tree diagrams). For clarity, we have displayed the affricates as contoured segments, i.e., with one root node combining two ‘contour’ nodes, although we note that in some accounts (including Clements' feature geometry) affricates are non-contoured segments, with the release captured by the simple feature [strident] (Clements, 1999), or in the case of lateral affricates, by the feature [lateral] (Chomsky & Halle, 1968: SPE). For present purposes we designated the Zulu fricatives, and the fricated release of the Tlingit affricates, as [-sonorant] regardless of their voicing, in keeping with traditional feature descriptions of fricatives more generally. It must be noted, however, that the role of [sonorant] in the feature geometry of lateral fricatives (and by implication, affricates), as compared to lateral approximants, has been under debate with respect to several other languages (e.g., Ball, 1990;Brown, 1995;Holton, 2001).

Given strong additional interests in cross-language voicing differences, we also addressed the effects of language experience on perception of feature- or gesture-relevant aspects of voicing by comparing native listeners of English and French, for which initial stop voicing differs both featurally and gesturally. Importantly, all the onsets under scrutiny are nonnative for both groups of listeners. Their perception of the three types of contrast could thus reflect language-specific phonotactic differences in permissible onsets. Whereas both English and French disallow word-initial /tl/ or /dl/, French allows all voiced fricative onsets but English disallows /ʒ/ as an onset. Conversely, English allows affricates but French does not, except in loanwords (e.g., “jeep” or “check-up”). Still, listeners' perception of voicing may simply reflect English-French differences in how voicing is phonetically implemented in the two languages. The gestural approach claims that obstruent voicing distinctions arise from presence/absence of a glottal opening-closing gesture in the two languages. But English and French phase the glottal gesture differently relative to supralaryngeal constriction gestures for voiceless obstruents

[9]In featural terms, it is an aspiration contrast ([Ø]-[spread glottis], rather than [±voice]). The only voiced consonants of Tlingit are sonorants: a single nasal (/n/) and two glides (/j, w/). For all obstruents, Tlingit has a three-way laryngeal contrast among (voiceless) unaspirated ([Ø]) versus aspirated ([spread glottis]) versus ejective ([constricted glottis]) (Maddieson et al., 2001). Thus, there appears to be no independent [voice] feature in Tlingit.

(onset vs. offset of the constriction), and relative to the timing of the glottal adduction that permits glottal pulsing for voiced obstruents (see footnote 5).

The phonological feature descriptions are less straightforward. The feature geometry description is that English uses [spread glottis] for voiceless obstruents (fricatives as well as stops), regardless of utterance position, whereas French uses [±voice] to distinguish voicing.[10] By the gestural account of articulatory phonology, however, presence/absence of a glottal opening-closing gesture determines voicelessness or voicedness, respectively, for both languages. In the absence of this active laryngeal gesture, the default glottal configuration for speaking is a"critical" degree of vocal fold constriction, i.e., just-sufficient adduction of the folds to allow glottal pulsing under appropriate conditions of subglottal pressure, vocal fold tension, and airflow (e.g., Goldstein & Browman, 1986). English and French are thought to differ gesturally, in stress-initial position, with respect to the degree of glottal adduction used in resting position; this results in differences in glottal pulsing during oral closure for these two languages.

What, then, are the possible implications of these differences and commonalities between English and French for nonnative voicing perception? Phonetically, the coronal stops in the Hebrew contrast /dl/-/tl/ represent a better approximation of French than English stop voicing contrasts. In featural terms, [-voice] vs. [+voice] can be extracted from the surface of Hebrew /dl/-/tl/, thus fitting the French system but not the English system of Ø-[spread glottis]. The opposite situation holds for Tlingit /dɮ/-/tɬ/, from which only Ø vs. [spread glottis] can be extracted, thus fitting the English but not the French system. In gestural terms, both the Hebrew and Tlingit voicing contrasts involve presence/absence of a glottal opening-closing gesture. However, in Hebrew /tl/ the glottal gesture is in-phase with oral closure, as in French voiceless initial stops, whereas in Tlingit /tɬ/ ([t͡ɬʰ]) it is presumably phased with closure release (see footnote 9), as in English voiceless initial stops (see footnote 5).

What predictions might the two accounts offer regarding listener language differences in perception of voicing across the target contrasts? On the gestural account, English and French listeners should both perform well on perception of voicing in Hebrew /dl/-/tl/, which is distinguished by the presence/absence of a glottal opening-closing gesture, as are initial stop voicing contrasts in both listener languages. However, because that gesture is phased to oral *closure* in the voiceless stop, more akin to the French intergestural organization for voiceless stops, French listeners may perform somewhat better than English listeners *if the* full gestural structure of the onset affects voicing perception. Conversely, French listeners should do worse than English listeners with the Tlingit affricate voicing distinction, in which the glottal opening-closing gesture is phased with oral release rather than closure, more akin to the intergestural organization for English voiceless stops. The featural account suggests that English listeners should have serious difficulty with the Hebrew /dl/-/tl/ [voice] contrast, which is not distinctive in English, and that French listeners should have great difficulty with the Tlingit /dɮ/-/tɬ/ Ø-[spread glottis] contrast, which is not distinctive in French. As for Zulu /ɮ/-/ɬ/, unequivocal predictions can be made by the gestural account: /ɬ/ but not /ɮ/ is produced with a glottal opening-closing gesture, peak-aligned with the oral constriction gestures. English and French listeners should be equally and highly sensitive to the presence/absence of that gesture in

[10]In English and other languages that contrast short- and long-lag stops stress-initially, phonologically voiceless stops are specified with the privative feature [spread glottis] in the feature geometry approach of Clements (2003). If [spread glottis] is unspecified (Ø) the stop is phonologically *voiced*. By contrast, languages that distinguish short-lag from pre-voiced stops (e.g., Spanish and French) use the binary feature [±voice] to indicate phonological voicing. Their phonologically voiceless [-voice] stops, usually realized as short-lag unaspirated, may be *phonetically* similar to phonologically *voiced* English initial stops. Moreover, by the feature economy principle of 'mutual attraction' (Clements, 2003), languages with voicing contrasts in more than one obstruent type should tend to apply the same phonological feature to all. Fricative voicing would thus be [spread glottis] vs. Ø for English, but [±voice] for French). Some scholars have proposed a universal *phonological* distinction of [±voice] (Keating, 1990) or [±spread glottis] (Vaux, 1998) to clarify this confusing situation.

fricatives. The featural account, however, makes differential predictions for the two listener groups: Zulu /ɮ/-/ɬ/ is a [±voice] contrast, a distinction consistent with French but not English obstruent voicing. Thus from a featural viewpoint, /ɮ/-/ ɬ/ should be "easy" for French but difficult for English listeners.

Turning now to the primary issue that motivated this study, we ask: What might the contextual dependency differences among these onsets imply for nonnative voicing perception? We have identified two *intergestural* properties of onsets that could influence perception of initial voicing distinctions (see Table 1 and Figure 1): overall gestural complexity and tightness of intergestural phasing. We use complexity to refer to the number of gestural specifications required to capture the onset's structure. Given that the features of a phonological segment take on independent values, onset complexity should be irrelevant, by a featural account, to the function of any single feature. As for gestures, gestural tightness refers to how strict the phasing, or coupling, is among the laryngeal and supralaryngeal gestures: are they all in-phase together (tight) or are some anti-phased or in some other phasing relationship, such as peak-phased (looser coupling)? This factor is also unaddressed (and irrelevant) in featural accounts, in which features are static, timeless properties of phonological segments. By the gestural account, one or both types of gestural interdependency should affect voicing perception, as we previously argued. In contrast, the featural account predicts minimal to no impact of either type of contextual dependency, especially if features are independent. Note that these contextual effects should apply across the two listener languages, above and beyond any language-specific effects in voicing *per se*.

2. Experiment

2.1. Method

2.1.1. Participants—Native speakers of American English (*n* = 19) and Parisian French (*n* = 16) (henceforth English and French listeners) participated for course credit or participant fee. None had been exposed to the languages or onset contrasts tested. None reported hearing, speech or language problems. Twelve additional English listeners were tested but their data were not retained due to inappropriate linguistic background, failure to complete all test sessions, and/or a high missing response rate (> 2.5 s.d. above the group mean).

2.1.2. Stimulus Materials—Male native speakers[11] of Hebrew (born and raised in Tiberius, Israel), Tlingit (Juneau, Alaska), and Zulu (Durban, South Africa) recorded 20 repetitions each of open syllables with the targeted onsets of their native language followed by /a i u/; only the /a/ context was used for this study. Stimuli were prompted individually in print with real-word examples (e.g., Hebrew: "DLA as in [12]דלעה"), in randomized order, and recorded in a sound-attenuated room with a Shure SM10A dynamic headset microphone and Sony portable DAT recorder. For the final stimulus set, five tokens each of the two members of each contrast were chosen for similarity in F0 level/contour and duration. In each target contrast the voicing-related differences were consistent with prior reports on voicing distinctions in stops and fricatives, according to detailed acoustic measurements (duration, integrated energy and spectral peaks of the burst/fricative; VOT for stops/affricates; F0 and formant frequencies at three points in the vowel, as well as in the lateral approximant [Hebrew] or frication [Tlingit, Zulu]; segment durations); the selected tokens within each contrast had comparable values on the other, voicing-irrelevant dimensions. In Hebrew /tl-/dl/, the /t/ had a relatively short voicing lag and the /d/ was pre-voiced (+38 vs. -148 ms VOT); integrated energy over stop release was greater for /t/ than /d/ (3.8 vs. 1.3 dB); and F0 at /l/ onset was higher for /tl/ than /dl/ (134 vs.

[11]All were fluent L2 speakers of English.
[12]*dla 'at* ('pumpkin')

114 Hz) (all similar to those in Hallé and Best, 2007). In Tlingit, both / tɬ/ and /dɮ/ showed voicing lag; the initial voiceless portion of the lateral release was longer for /tɬ/ than /dɮ (147 vs. 38 ms) and the release was longer overall for / tɬ/ than /dɮ/ (202 vs. 139 ms), thus 26% of the release was voiced in / tɬ/ vs. 73% in /dɮ/; integrated energy over the voiceless portion was greater for /tɬ/ than /dɮ/ (9.3 vs. 2.5 dB); F0 at onset of /a/ was higher in /tɬa/ than /dɮa/ (142 vs. 128 Hz). Zulu /ɮ/ was fully voiced and ɬ/ was almost totally voiceless; frication was slightly longer for /ɮ/ than /ɬ/ (221 vs. 203 ms); F0 was higher at /a/ onset in ɬ/ than /ɮa/ (112 vs. 96 Hz). Thus, the Hebrew stop contrast was prevoiced versus voiceless, for Tlingit affricates it was short-lag versus long-lag voiceless, and for Zulu fricatives it was simply voiced-voiceless. For illustrative spectrograms and waveforms, see Figure 3.

2.1.3. Procedure—The experiment was run in a sound-proof booth, using a Macintosh Powerbook G4 with Psyscope software. Participants took discrimination and categorization tests on each contrast. Discrimination was tested first, to minimize effects of stimulus categorization on performance. A categorial AXB task was used, given its fairly low memory demand and low response bias. Open-response categorization was used to reveal how listeners' assimilations to their native language reflected their perception of the differing structures of the nonnative onsets. Closed-response tasks are more typically used in cross-language perception studies, but were deemed to be too likely to constrain the listeners' responses to the experimenters' own intuitions about possible percepts.

AXB Discrimination: For each contrast, there were 60 AXB triplets: 15 different token pairings, all presented in each of the 4 triad types (AAB, ABB, BBA, BAA), with the constraint that each of the five selected syllable tokens for each member of the contrast appeared equiprobably in each position. A and B always differed with respect to voicing and the target item X was always a different token than the A or B item that it matched phonologically. Trials were presented in random order, blocked by language; block order was counterbalanced across subjects (interstimulus interval = 1 s, intertrial interval = 3 s, interblock interval = 5 s). There was a 5-trial training phase. On each AXB trial, participants indicated whether X categorially matched A or B by pressing buttons labeled ‘1’ and ‘3’. They were told to respond on each trial, even if guessing, and to respond as quickly as possible after hearing the third item. Missed trials (response time > 2 s) were recycled so that each subject completed all 180 trials.

Categorization: Each of the 30 stimulus tokens was presented 3 times, in random order, for 90 trials. This was preceded by a training phase of 6 trials. On each trial, participants were presented with a syllable, which they had to transcribe using the keyboard. If they were ambivalent re: several transcriptions, they could report them all, using ‘/’ to separate them.

2.2. Results

2.2.1. Discrimination—Figure 4 shows the discrimination performance for accuracy (A) and response time (RT) data (B). Analyses of variance were conducted on the percent correct and on the response time data, on the within-subject factors *Target* (voiced vs. voiceless X in AXB trials), triad *Pattern* (XXY vs. XYY trials, respectively referred to as primacy vs. recency) and stimulus *Language* (Hebrew, Tlingit, Zulu), and the between-subject factor listener *Group* (English vs. French). Target was significant in both the accuracy and the RT data, $F(1,33) = 4.38$; $p < .05$, but did not interact with Group or Language: discrimination was better when the target (X) was voiced than voiceless. Pattern was only marginally significant and did not interact with Group or Language.

Of greater interest are the experimental factors. We report first the statistics on the accuracy data. Group was significant, $F(1,33) = 13.81$, $p < .001$, reflecting an overall advantage for English over French listeners (93.9% vs. 88.6% correct). However, the significant Language

× Group interaction, $F(2,66)= 53.69$, $p < .05$, indicates that this advantage differed according to stimulus Language. English listeners performed better than French listeners only for the Tlingit contrast (90.1% vs. 69.8% correct), $F(1,33) = 43.02$, $p < .0001$ (see Figure 4). On the two other contrasts, both French and English listeners performed near ceiling, with a tiny advantage for French over English listeners. This advantage was nonetheless significant for Hebrew (99.0% vs. 97.1%), $F(1,33) = 4.47$, $p < .05$, but marginal for Zulu (97.0 % vs. 94.5%), $F(1,33) = 3.04$, $p > .05$. The Language main effect, $F(2,66) = 110.94$, $p < .0001$, indicates that performance for Tlingit was significantly lower than for Hebrew, $F(1,33) = 129.01$, $p < .0001$, and Zulu, $F(1,33) = 127.46$, $p < .0001$. Performance was also lower for Zulu than Hebrew, $F(1,33) = 8.07$, $p < .01$, that is, Hebrew > Zulu > Tlingit.

The response time data paralleled, in reverse, the correct discrimination rate data, with longer mean RTs for the lower mean discrimination rates. English listeners were on average faster than French listeners on the Tlingit contrast (1365 ms vs. 1467 ms), but slower on the Hebrew (869 ms vs. 750 ms) and Zulu (1009 ms vs. 878 ms) contrasts. These differences, although numerically large (mean differences > 100 ms), did not reach statistical significance due to large inter-subject variability. However, the Language × Group interaction was marginally significant, $F(2,66) = 13.47$, p = .071, mirroring the differential pattern of performance found in the accuracy data.

2.2.2. Categorization—The categorization data provide complementary insights about the perceived structures of the nonnative onsets. Participants' open-response spellings ("naïve transcriptions") of the target onsets were re-coded according to perceived voicing, manner, place of the "primary," and "secondary" consonants transcribed (e.g., /k/ and /l/ respectively, for transcription of /tl/ as /kl/), and onset complexity (i.e., singleton vs. multi-segment) by two phonetically trained late bilinguals (CTB: Eng-Fr; PAH: Fr-Eng). Analyses of variance (ANOVAs) were conducted on the dependent measures of perceived voicing, place, and complexity. In each ANOVA, stimulus *Language* (Hebrew, Zulu, Tlingit) and *Target* voicing (voiced vs. voiceless) were within-subject factors, listener *Group* (American English vs. French) was between-subject.

2.2.2.1 Perceived voicing: The significant Language effect, $F(2,66) = 26.43$, $p < .0001$, was compatible with the discrimination findings: Voicing identification was best overall for Hebrew clusters (95% correct), worst for Tlingit lateral affricates (77%), and intermediate for Zulu lateral fricatives (87%). Simple effect tests of a Language × Group interaction, $F(2,66) = 37.74$, $p < .0001$, found that voicing was more accurately identified by French than English listeners for both Hebrew clusters (French: 100% correct; English: 92%), $F(1, 33) = 5.37$, $p < .05$, and Zulu fricatives (French: 95%; English 80%), $F(1,33) = 6.86$, $p < .05$, but this was dramatically reversed for the Tlingit affricates (English: 91%; French: 61%), $F(1, 33) = 56.26$, $p < .0001$. Overall, perceived voicing accuracy was lower for voiced (79%) than voiceless (94%) onsets, Target: $F(1,33) = 58.18$, $p < .0001$, but the difference was much larger for French (voiced: 72%; voiceless: 99%) than English listeners (voiced: 84%; voiceless: 90%), Target × Group: $F(1,33) = 58.18$, $p < .0001$. The primary source of the significant Language × Target, $F(2, 66) = 42.55$, $p < .0001$, and Language × Target × Group interactions, $F(2, 66) = 17.63$, $p < .0001$, was traced to Tlingit, which alone produced significant simple effects of Target (voiceless: 99% correct; voiced: 55%), $F(1, 33) = 153.21$, $p < .0001$, and Group × Target, $F(1, 33) = 63.26$, $p < .0001$. The groups did not differ on voiceless targets (English: 99%; French: 99.6%), but the English out-performed the French on voiced targets (82% vs. 23%).

2.2.2.2 Perceived place of articulation: We used percent of transcriptions that included velars (/k, g/) to examine coronal-to-velar place shifts in perception of the target onsets. The velar shift was larger overall for English (67% velars) than French listeners (59%), Group: $F(1, 33) = 5.92$, $p < .05$. The velar shift was nearly at ceiling for Tlingit (98%), strong but lower for

Hebrew (78%) and rather infrequent for Zulu (14%), Language: $F(2, 66) = 206.30$, $p < .0001$. Simple effects tests show that for Hebrew onsets, more velars were reported by English (86%) than French listeners (69%), Group: $F(1, 33) = 4.75$, $p < .05$, whereas the groups failed to differ on either Zulu, where the velar shift was quite weak (English: 17.5%; French: 9%), or Tlingit, where the velar shift was conversely at ceiling (English: 97%; French: 99%).

Overall, more velars were reported for voiceless (67%) than voiced onsets (60%), Target: $F(1, 33) = 6.58$, $p < .05$, but this was modulated by interactions of Target × Group: $F(1, 33) = 20.16$, $p < .0001$, Language × Target: $F(2, 66) = 27.82$, $p < .0001$, and Language × Target × Group: $F(2, 66) = 4.35$, $p < .05$. Velar percepts were more frequent for voiceless (98%) than voiced (59%) Hebrew onsets, Target simple effect: $F(1, 33) = 36.44$, $p < .0001$, but the difference was much larger for French (voiceless: 98%; voiced: 35%) than English listeners (voiceless: 92%; voiced: 72%), Target × Group: $F(1, 33) = 36.44$, $p < .0001$. For the much weaker velar shift with Zulu, the asymmety was reversed -- more velars were reported for voiced (22%) than voiceless lateral fricatives (5%), Target: $F(1, 33) = 5.44$, $p < .05$. This pattern is attributable to the English listeners (voiced: 35%; voiceless: 0%); the French shift was smaller and the asymmetry was in the opposite direction (voiced: 7%; voiceless: 11%), Target × Group: $F(1, 33) = 8.78$, $p < .01$. Target and Group failed to affect the Tlingit velar shift.

2.2.2.3 Perceived onset complexity: Complex onsets, defined as percent of onsets heard as sequences of two or more consonants,[13] were overwhelmingly reported for Hebrew (95%) and lower but still a substantial majority for Tlingit (79%), while Zulu onsets were much less often heard as complex (32%), Language: $F(2, 66) = 119.64$, $p < .0001$. This pattern was qualified by interactions of Language × Group, $F(2, 66) = 6.44$, $p < .005$, Language × Target Target, $F(2, 66) = 3.20$, $p < .05$, and Language × Target Target × Group, $F(2, 66) = 12.47$, $p < .0001$, which were examined by simple effects ANOVAs. For Hebrew, both groups nearly always reported complex onsets (French: 99.8%; English: 96%; usually stop +/l/), though the tiny difference was significant, Group: $F(1, 33) = 4.65$, $p < .05$. The Zulu lateral fricatives were heard as complex onsets much less often than Hebrew onsets; also, mostly heterorganic stop +fricative were reported, never stop+/l/. French listeners reported complex onsets more often for voiceless than voiced Zulu fricatives (voiceless: 54%; voiced: 30%); English listeners showed the opposite asymmetry (voiced: 35%; voiceless: 14%), Group × Target: $F(1, 33) = 11.00$, $p < .05$. The most frequent *singleton* answers were fricatives, both for English (voiced: 50%; voiceless: 74%) and French listeners (voiced: 53%; voiceless: 43%); stops and affricates were rarely reported (English; 13.5%; French: 10%). The Tlingit lateral affricates elicited intermediate reports of complex onsets, which were more frequent for English (87%) than French listeners (70%), and for voiced (90%) than voiceless onsets (69%). The two groups reported complex onsets equally often for voiced affricates (English: 92%; French: 87%), but French listeners heard fewer than English listeners for voiceless affricates (54% vs. 81%), Group × Target: $F(1, 33) = 16.11$, $p < .0005$.[14] Voiceless /tɬ/ elicited more stop+/l/ from English than French listeners (78% vs. 54%, $p < .0001$, *t*-test). French listeners reported singleton stops fairly often for /tɬ/ (46%; English: 18%, $p < .0001$, *t*-test), but not for /dʒ/ (French: 13%; English: 7%, *ns*).

3. Discussion

Discrimination accuracy and correct-response reaction time (RT) converge to show that, for all listeners, the Tlingit /dʒ/-/tɬ/ voicing contrast was more difficult than those of Zulu and Hebrew. English listeners also performed slightly but significantly better overall than French

[13]Here, we treated affricates as singletons, given the general phonological premise that they are single segments.

[14]Interestingly, Belgian French listeners perceive the stops of Quebecois French (Canada) preceding high front vowels, where they are affricated, as stop-fricative clusters (Béland & Kolinsky, 2005).

listeners. However, the greater interest is in the interactions: English listeners strongly outperformed French listeners on the Tlingit contrast (about 20% higher accuracy and 100 ms faster RTs), which is congruent with the known difference in phonetic settings for voicing between English and French initial stops. By comparison, French listeners were slightly better, and notably faster, than English listeners on the Zulu and Hebrew contrasts (about 2-3% higher accuracy and > 100 ms faster RTs, though inter-subject variability kept these differences from achieving significance), on which both groups' discrimination accuracy approached ceiling.

The categorization data obtained from open-responses were closely congruent with the discrimination data. Indeed, correct discrimination rate was predictable from the full assimilation patterns of the open-response categorizations to English/French or, for the most part, from just the voiced/voiceless responses implicit from the listeners' categorizations. The poor discrimination performance of the French listeners on Tlingit /dɮ/-/tɬ/ could thus be predicted by their frequent categorization of *either* /tɬ/ or /dɮ/ as voiceless.[15] The open response categorization data also revealed three interesting patterns: (1) Zulu lateral fricatives were often analyzed by the listeners as complex onsets with a lateral component (English listeners) or with a fricative component (French listeners); conversely, both groups showed a trend toward analyzing Tlingit affricates, but not Hebrew clusters, as singleton stops; (2) the coronal place of the Tlingit affricates and of the stops in the Hebrew clusters were very often misperceived as velar; (3) there was an asymmetry in the coronal-to-velar effect between Hebrew /dl/ and /tl/, especially marked for French listeners, with voiceless /tl/ inducing much more velar perceptual shifts than did voiced /dl/. The latter result is consistent with earlier findings on French and English listeners' perception of /dl, tl/, which are phonotactically impermissible onsets in both languages (Hallé & Best, 2007;Hallé et al., 1998), and thus confirms the robustness of this case of “phonotactic assimilation.”

Two primary conclusions can be drawn, moreover, regarding the primary issue of interest, the perception of voicing contrasts in nonnative onset types. First, perception of voicing in the three nonnative target onsets differs according to the listeners' native voicing distinctions. English and French listeners most clearly differed in their performance on the Tlingit lateral affricates. This is consistent with classic and more recent reports on native-language influences in perception of nonnative voicing contrasts, and extends them to three new types of nonnative syllable onsets.

The more novel conclusion is the second one: perception of voicing varies with the structure of the nonnative onsets. For both groups, discrimination and identification of the Hebrew clusters was better and more consistent than for the Zulu fricatives, which in turn was better and more consistent than for the Tlingit affricates. This pattern is most easily interpretable on an articulatory phonology analysis, specifically in terms of gestural phasing, or tightness of gestural coupling, in the three types of onsets. Listeners found it more difficult to perceive voicing as a separate parameter the more tightly coupled the glottal and oral gestures were (refer to Table 1). They appear to have perceived the onsets as complex gestural structures including voicing, place, manner, and often as involving multiple segments even when the target item was mono-segmental.

Overall, the findings are consistent with the view that onsets serve as coherent units of phonological structure, and that their gestural organization, particularly their pattern of intergestural phasing, is critical to the perception of phonetic distinctions, here the voicing

[15]We computed discrimination performance from categorization data for each subject, including in the calculation *any* perceived phonological differences between the members of a contrast. Predicted and actual performance were found to be very close (mean difference 2.5%) and followed the same patterns. A lesser fit obtained when just perceived voicing differences were included but, again, followed the same patterns of performance.

distinction. In the following, we address successively the two classes of predictions that were made in the introduction: Language-specific effects on perception of voicing, and the influence of the gestural structure of syllable onsets on perception of voicing distinctions.

3.1. Listener language effects on perception of nonnative voicing contrasts

English and French listeners differed most notably on Tlingit /dɮ/-/tɬ/, a voicing contrast that is phonetically similar to English but not French initial stop voicing contrasts. In terms of feature description, the Tlingit contrast fully matches the English voicing contrast under our current assumption of [spread glottis] as defining "voiceless" in English but not in French, where the distinction is instead [±voice]. That listeners perceive nonnative distinctions in terms of the phonetic settings of their native phonological contrasts is not in itself a novel finding. However, the present results do offer novel insights in their extension of perceptual investigations of voicing contrasts beyond simple stops to nonnative affricates, fricatives and stop+liquid clusters. Consistent with both the featural and gestural account predictions, English listeners discriminated Tlingit /dɮ/-/tɬ/ better than French listeners, most likely for two reasons. First, affricates are foreign to French phonology, and their relatively unfamiliar gestural organization also contributed to perceptual difficulty for French listeners, as evident in their frequent transcription of /tɮ/ as a simple stop (46%). Second, the voicing contrast for the Tlingit affricates is phonetically comparable to that of English stops, including affricates, but is unlike the voicing distinction in French stops. Compatibly, whereas English listeners "correctly" perceived Tlingit /dɮ/ as voiced 79% of the time, French listeners instead perceived it as voiceless the majority of the time (65%).

These results, however, do not reflect a perfect match between the English and Tlingit phonological contrasts nor a complete mismatch between French and Tlingit, as the feature account would predict. Specifically, French listeners reported Tlingit /dɮ/ as voiced (22%) or voice-ambiguous (11%) substantially often, even though the /dɮ/ tokens were voiceless unaspirated, compatible with French voiceless stops. As for English listeners, they misperceived Tlingit /dɮ/'s voicing to a non-negligible degree (21%) despite its phonetic comparability to English phonologically voiced stops and affricates. These observations are not surprising from the gestural perspective, which assumes that both the English and the French voicing contrasts are determined by the presence/absence of a glottal opening-closing gesture, and that this gesture appears as part of the phased intergestural complex that comprises the syllable onset, where the languages differ in whether they phase the glottal gesture to closure (French) or release (English), which has differential acoustic consequences. The Tlingit /dɮ/-/tɬ/ contrast is a voiceless unaspirated-aspirated contrast, which we assume to be produced with a clear glottal opening-closing gesture for /tɬ/ but not for /dɮ/, where glottal adduction to produce glottal pulsing is apparently phased with the release of supralaryngeal closure (no pre-voicing) rather than prior to supralaryngeal closure (as with French pre-voiced stops). The Tlingit voicing contrast thus fits better with the English than the French pattern of stop voicing.

For the other two contrasts, the groups differed in discrimination accuracy, and differed substantially though not significantly in RT. They also differed in certain aspects of their naive transcriptions, including not only their categorizations of voicing, but also their categorizations of the place and phonological complexity of the target items. For Hebrew /dl/-/tl/, the equivalent discrimination performance and voicing categorization of English and French listeners is contrary to the predictions of the featural account but fully compatible with those of the gestural account. That the French discrimination and voicing categorization was excellent on Hebrew /dl/-/tl/ can be explained by the perfect match between French and Hebrew for either the feature or the gesture characterizations of native voicing contrasts. However, for English this is not the case; the featural account fits notably less well than the gestural account. English and Hebrew do not correspond either in terms of features (Ø-[spread glottis] vs. [±voice]) or surface

phonetic-acoustic properties (short- vs. long-lag VOT as compared to negative vs. short-lag VOT). So why did the English listeners perform near ceiling on the Hebrew voicing contrast? One important reminder is that the earlier-reported difficulties that listeners of English have had with nonnative prevoiced versus short lag unaspirated stop voicing contrasts (Abramson & Lisker, 1970) were restricted to simple singleton stops, whereas the present Hebrew stimuli were clusters rather than singletons. While the very good discrimination and voicing categorization performance of English listeners on Hebrew /dl/-/tl/ are surprising from a featural or acoustic point of view, they are not so surprising from a gestural point of view, in that the stops in the Hebrew clusters contrast on presence/absence of a glottal opening-closing gesture.

Further support for a gestural contribution to the perception of voicing may be inferred from earlier reports on perception of nonnative stop voicing contrasts. First, English listeners *have* shown excellent discrimination of certain specific classes of prevoiced versus short-lag unaspirated nonnative stops: the clicks of Zulu, which have a three-way voicing distinction among pre-voiced, short lag unaspirated, and long lag aspirated (Best, McRoberts & Sithole, 1988; Best & Avery, 1999). In contrast, English listeners performed poorly on Zulu/Xhosa pre-voiced implosive versus voiceless short lag stops /ɓ/-/p/ (Best et al., 2001), whereas Spanish-English bilinguals (with English as L2) performed well on the same contrast (Calderon & Best, 1996). This cross-language difference fits well with the classic VOT perception data, as /ɓ/-/p/ is acoustically similar to the Spanish prevoiced vs. voiceless unaspirated contrast /b/-/p/, but both consonants are similar to English initial /b/ (which may be either voiceless unaspirated or prevoiced). Thus, the data on the perception of nonnative prevoiced versus short lag contrasts do not uniformly reveal perceptual difficulties in English listeners. As Best et al. (1988) proposed, the key to the good performance of English listeners in the click case may be that they perceived the clicks as nonspeech, rather than as native consonants. Clicks fit the PAM's “nonassimilable” (NA) category for speakers of non-click languages, such that English listeners fail to perceive Zulu clicks as speech sounds (see Best & Avery, 1999, for functional evidence). Anecdotal observation suggests that auditory streaming takes place when non-click language listeners are confronted with connected speech containing click consonants. That is, the noise bursts at click release make a separate perceptual stream from the remaining consonants and vowels, which could be experienced by English listeners as a string of separate nonspeech sounds. More research is of course needed to confirm and understand this perceptual pattern, and its contribution to perception of nonnative prevoiced-voiceless unaspirated contrasts by listeners of languages like English. At this point, we may only speculate that the streaming illusion may help in factoring out those speech attributes that are critical to voicing perception, thus reducing the effects of contextual complexity on perception of nonnative syllable onsets.

The Hebrew /dl/-/tl/ contrast provides a novel case in which English listeners are good at discriminating a prevoiced versus short lag voicing contrast, which is unexpected according to prior phonetic feature analyses. It is unclear at this point whether or how the speculative account offered for the clicks may apply to English listeners' perception of voicing in the Hebrew clusters. Alternatively, the very good English performance on Hebrew /dl/-/tl/ might be due to a peculiarity of this contrast that some have reported: whereas Hebrew voiced stops are uncontroversially prevoiced with a substantial voicing lead, Hebrew voiceless stops have been described as “medium-lag” VOT (Raphael, Tobin, Faber, Most, Kollia, & Milstein, 1995). If this is correct, then [spread glottis] might be extracted from Hebrew /tl/ by English listeners. The /t/s in the Hebrew /tla/s used in the current experiment had a mean VOT of 38 ms, a value that can hardly be viewed as long or even medium voicing lag as in English; it is probably best considered short-lag. Still, the corresponding delay of the glottal opening-closing gesture may be sufficient to be perceived as phased with the supralaryngeal release (as in English voiceless stops) rather than with the closure (as in French short-lag voiceless stops).

Finally, English and French listeners's performance on Zulu /ɮ/-/ɬ/ did not differ statistically in discrimination accuracy or response time, but French listeners were better than English listeners on voicing categorization (95% vs. 79% correct voicing). This result is quite in line with Best et al.'s (2001) data for the Zulu /ɮɛ/-/ɬɛ/ contrast, which found near-ceiling discrimination by English listeners but only 82% correct voicing categorization. The slight advantage we found for French over English listeners in categorizing the voicing of Zulu /ɮ/-/ɬ/ is difficult to interpret from either the gestural or the featural approaches. Gesturally, /ɮ/-/ɬ/is determined by absence versus presence of a peak-phased glottal opening-closing gesture to which French and English listeners should be, in principle, equally sensitive. In phonetic feature terms, if [±voice] rather than [spread glottis] is readily extracted from Zulu /ɮ/-/ɬ/, this could explain the slight difficulty in voicing categorization encountered by English listeners, and their non-significantly slower discrimination responses, but it could not explain their near-ceiling discrimination performance.

To summarize, then, the language-specific effects on perception of voicing in the three nonnative onset contrasts are compatible in some ways with both featural and gestural viewpoints, though somewhat less than perfectly for either perspective. However, considering all three nonnative contrasts, overall the differences between the French and English listener results appear to fit the gestural account better than the featural account.

3.2. Effects of stimulus language differences in syllable onset structure

The second, and more novel conclusion we can draw from the study is that there are systematic effects of syllable onset structure on perception of voicing distinctions across the listener groups. For both groups, discrimination of the Hebrew clusters was slightly but significantly better than for the Zulu fricatives, which was in turn much better than for the Tlingit affricates. The voicing categorization data followed the same pattern. We argue that these results reveal a systematic effect of the tightness of intergestural phasing on speech perception. More specifically, voicing perception is increasingly influenced by increasingly tighter phasings between laryngeal and supralaryngeal gestures. This is clearest for the English performance. Based on featural analysis or on surface phonetic-acoustic form, English listeners should have performed better on Tlingit /dɮ/-/tɬ/ than on Hebrew /dl/-/tl/, but the opposite result was found.

How would the gestural account accommodate this finding? We propose that the perceptual pattern we observed is attributable to an increased difficulty imposed by the gestural organization of the Tlingit onsets relative to that of the Hebrew onsets. As described earlier (also see Figure 1A), the /d, t/ and the /l/ gestures in Hebrew /dl/-/tl/ clusters are anti-phased; the glottal gesture for /t/ is tightly linked (in-phase) only to the oral stop gesture and not to the lateral gesture. By comparison, in Tlingit, all the gestures for oral stop and lateral release are tightly phased both with one another and with the glottal gesture (all are in-phase: see Figure 1C). The larger perceptual difficulty thus coincides with an overall tighter gestural coupling/phasing specification. The pattern of French performance would not be sufficient in itself to test for the influence of this phasing specification on perception because French listeners, unlike English listeners, should perform better on prevoiced versus short lag, than on short versus long lag anyway. But the dramatic decline in French performance from Hebrew to Tlingit (e.g., 99% to 70% for correct discrimination) is probably not fully accounted for by the phonetic mismatch alone. We thus propose that both French and English listeners found it more difficult to perceive voicing as a separate parameter, the more tightly coupled the glottal and oral gestures of the nonnative syllable onset were.

That proposal is also compatible with the somewhat lower performance of both French and English listeners on Zulu /ɮ/-/ɬ/ than Hebrew /dl/-/tl/. In the Zulu voiceless fricatives, as in most voiceless fricatives, the peak of a glottal opening-closing gesture for the voiceless fricative is phased with the peak of the supralaryngeal constriction gestures, in this case tongue tip and

tongue body constrictions. Peak-to-peak phasing is not as tight a coupling constraint as onset-to-onset phasing (since gestural peaks are not as precisely definable as onsets are) but is constraining nonetheless. Moreover, in lateral fricatives, the tongue tip and tongue body constriction gestures are onset-to-onset phased, adding to the coupling constraints for /ɮ/-/ɬ/ (see Figure 1B) as compared to /dl/-/tl/. By the gestural account we propose, these coupling constraints, while not as extreme as in Tlingit /dɮ/-/tɬ/, add to the contextual complexity in both production and perception, which explains the slightly less accurate and slower discrimination for Zulu /ɮ/-/ɬ/ as compared to Hebrew /dl/-/tl/.

In the introduction, we reasoned that two types of articulatory complexity could play a role in perception: gestural complexity and intergestural phasing tightness, roughly laid out in Table 1 as a first approximation. Both contextual gestural factors do indeed seem to play a role. For similar phasings (Tlingit and Zulu), voicing is extracted less easily from the gesturally more complex segments (Tlingit > Zulu); and for equivalent gestural complexities (Hebrew and Tlingit), voicing is extracted less easily from onsets in which intergestural phasing is more tightly specified (Tlingit > Hebrew). Yet, as discussed earlier, for French listeners the difference between Tlingit and Zulu is complicated by the non-French phonetic setting (phasing of the glottal opening-closing gesture to supralaryngeal constrictions) for the Tlingit voicing contrast. For this reason, the English listeners' performance is more directly informative about the impact of articulatory complexities alone. For them, the difference in performance between Tlingit and Zulu is also evident, though it is less dramatic than in the French listeners. The differences in English listeners could be attributed to the tighter intergestural phasing in the Tlingit affricates, instead of their greater gestural complexity, relative to the Zulu fricatives. By this reasoning, then, intergestural phasing tightness would be the primary structural factor underlying the voicing perception patterns we found. This notion is difficult to capture within the formalization of current feature models. Therefore, in the following, we offer only a gestural account of the findings.

3.3. Perception of complex onsets

We have reasoned that Zulu fricatives are gesturally simpler than the other onsets examined in this study. They combine similar supralaryngeal gestures as in /l/, tongue tip and tongue body constrictions that result in lateral airflow along the sides of the tongue, but they lack the coronal stop constriction preceding the lateral constriction, as in the Hebrew and Tlingit onsets. Zulu lateral fricatives differ from the /l/s commonly encountered by English and French listeners (apical alveolar approximants) in that their degree of constriction is greater (“critical” in articulatory phonology formulation; see Appendix), causing turbulence in the lateral airflow. Therefore, they are similar in complexity to ordinary /l/s segmentally, at least in terms of their supralaryngeal component gestures. Yet we found that listeners often reported complex onsets for Zulu /ɮ/-/ɬ/ (primarily as stop+fricative) in their open response categorizations to English or French (on average, 33% of the time), or infrequently as singleton affricates (~12%). These complex responses seem to deviate from the gestural organization of Zulu lateral fricative onsets with respect to intergestural phasing: two gestures that were simultaneous in the stimulus were perceived as sequential. Such perceptual responses may appear to contradict our tentative proposal that onset overall gestural organization in general, and temporal organization in particular, play an important role in perception. Timing is indeed perceptually important, given that the tightest intergestural phasing relations (Tlingit /dɮ/-/tɬ/) impacted most strongly on the perception of voicing. But this does not necessarily entail that listeners' categorization responses faithfully reflect the intergestural organization of the stimulus targets. Our data for Zulu fricatives suggest that both English and French listeners heard the supralaryngeal constrictions of /ɮ/-/ɬ/ separately as an apical closure and a critical lateral constriction of the tongue body, yet misperceived their peak-to-peak phasing relationship as anti-phase rather than in-phase.

We propose that such misperception is yet another illustration of a native language bias. Our interpretation is that nonnative onset structures are perceptually assimilated to acceptable ones in the listener's language, which are as “close” to the original onset structures as possible, according to an as yet underdefined articulatory metric. Just like Hebrew /tl/ and Tlingit /tɬ/ are assimilated to /kl/, the gestural structure of Zulu /ɮ/ and /ɬ/ appear to be perceptually assimilated to native segments or clusters. The novelty with Zulu onsets is that the assimilations repair the time structure of /ɮ/ and /ɬ/, rather than the place of articulation, so that these onsets are acceptable in the listeners' native language: for example, Zulu /ɬ/ may be perceived as /tʃ/, an acceptable affricate onset composed of the same the gestural components as /ɬ/. Language-specific differences between English and French listeners' responses also suggest that their percepts were motivated by phonological acceptability in their respective languages. For example, French listeners reported twice as many heterorganic stop+fricative onsets as English listeners, mostly clusters such as /ps/ that are permissible onsets in French but not in English. French listeners also reported /ʒ/ more often than /z/ for Zulu /ɮ/ (37% vs. 6%), consistent with French phonotactic patterns and frequencies of occurrence, whereas English listeners mostly reported /z/ and avoided /ʒ/ (35% vs. 3%), consistent with the fact that English words cannot begin with /ʒ/.

Based on these observations and the reasoning we have presented, we thus propose that perceptual assimilation extends to accommodate the time structure of nonnative onsets in which the crucial aspect that departs from listeners' native phonology has to do with their intergestural phasing relations. This notion helps to resolve the apparent puzzle of nonnative onset perception, indicating that it appears to be simultaneously both analytic and holistic. We propose that onsets are immediately perceived as a coherent, even if complex, structure that is composed of articulatory gestures with specific phasing relations to one another. Consistent with the well-known influence of language experience on perception of nonnative contrasts, complex nonnative onsets will be perceived through the lens of native-language restrictions on gestural sequencing and relative phasing (e.g., /tɮ/ perceived as /kl/; /ɬ/ perceived as /ts/), a phenomenon that has been referred to as “perceptual repair.” The implied “repair” does not necessarily entail consciously controlled processing. Indeed, several recent studies have found that phonological repair occurs very early in speech perception, at a prelexical level of processing (e.g., Dupoux et al., 2001; Hallé et al. 1998), and occurs outside of conscious control, for example, in the context of subliminal presentations (Hallé, Dominguez, Cuetos, & Segui, 2008).

Onsets thus appear to serve as coherent units of phonological structure, and their gestural organization figures critically in the perception of phonetic distinctions. Language-specific patterning of intergestural phasing relations is necessarily part of the perceived organization and may strongly influence perception of specific aspects of onsets. The present paper has focused on perception of voicing distinctions in nonnative onsets that differed in their intergestural structure, specifically, in the ways they coordinated the same subset of tongue tip, tongue body and glottal constriction gestures. As the categorization data suggest, the fact that listeners perceive onsets as coherent structures, including the phasing (timing) among gestures, also impacts on their perception of phonetic distinctions at the segmental level.

Acknowledgments

This work was supported by grants to the first author (NIH: DC00403) and to the second author (French Ministére de la Recherche: Cognitique LACO 1). We thank the following colleagues for their insightful comments on earlier drafts of the paper: Louis Goldstein, Michael Tyler, Ocke-Schwen Bohn, and Rachid Ridouane for his excellent guidance on feature theory and design of feature trees. We also wish to thank our two anonymous reviewers for their constructive suggestions on an earlier version of this paper. We are indebted and grateful to the native speakers of our target languages, who kindly advised us about their languages and made recordings for this work, as well as to our English and French listeners.

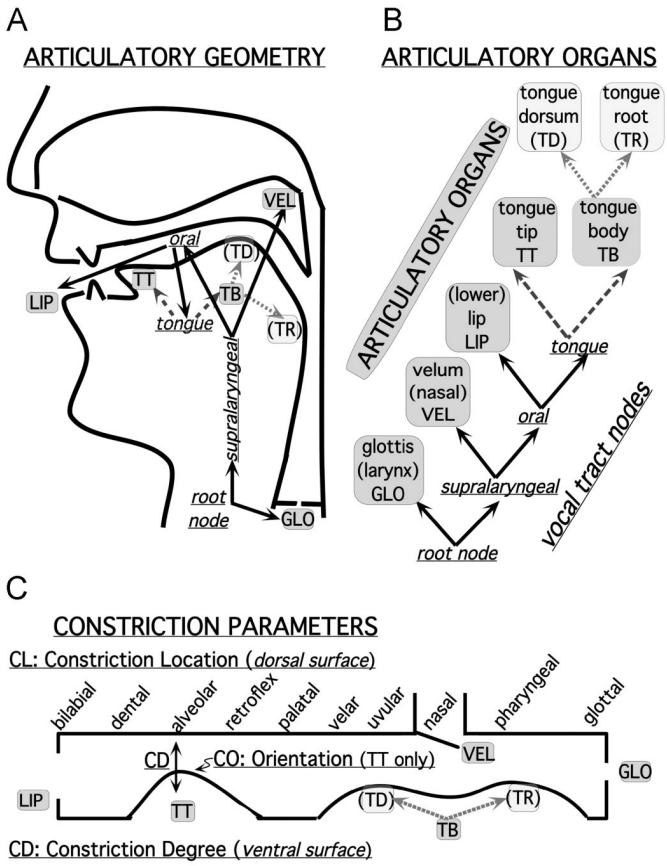
Appendix

NOTES ON ARTICULATORY GEOMETRY (Adapted from Browman & Goldstein's papers on Articulatory Phonology):

- Schematic of vocal tract tube geometry (A), straightened/flattened to display its articulatory tree structure (B)
- The active articulators are shown in the rounded boxes, as the end effectors of the articulatory branches:
  - GLO (glottis)
  - VEL (velum)
  - LIP (lower lip)
  - TT (tongue tip/blade)
  - TB (tongue body) may be further subdivided by some languages
    - TD (tongue dorsum)
    - TR (tongue root)
- The constriction parameters (C) are:
  - CL (constriction location, i.e., along the dorsal surface of the oral branch)
    - TT/TD/LIP: possible locations for a given end effector reflect both universal biomechanical constraints and language-specific phonotactic constraints (e.g., dental CL is possible for LIP, but alveolar CL is not)
    - LIP: lip protrusion (LP) is considered here to be CL = protruded lips
    - GLO: raised = implosive; lowered = ejective
  - CD (constriction degree):
    - closed (TT/TD/LIP: stops including affricates; VEL: non-nasal; GLO: ejectives, glottal stops)
    - critical (TT/TD/TR/LIP: fricatives; GLO: adduction for voicing)
    - narrow (TT/TD/TR/TR: approximants, high vowels; LIP: rounded vowels)
    - mid (TD/LIP: mid-height vowels)
    - wide (TD/TR/LIP: low vowels; VEL: nasals; GLO: opening gesture)

Further note that this geometry is used to denote individual articulatory gestures. The gestural scores for larger (higher) linguistic units, such as “segment,” syllable, word, or phrase, are comprised of task dynamic specifications of multiple articulatory gestures according to their stiffness, velocity, and phasing with respect to one another (i.e., in gestural constellations). To the extent that segments have linguistic or psychological reality, even at this level multiple gestures must often be phased with one another in constellations, e.g., GLO and an oral tract articulator are phased in language-specific ways to produce stop voicing contrasts, and

segments such as /r/ or /l/ involve multiple tongue articulators (also lips in some cases) whose gestures are phased to each other in language-specific ways.



References

Abramson AS, Lisker L. Discriminability along the voicing continuum: Cross-language tests. Proceedings, International Congress of Phonetic Sciences 1970;6:569–573.

Ball, MJ. The lateral fricative: lateral or fricative?. In: Ball, MJ.; Fife, J.; Poppe, E.; Rowland, J., editors. Celtic Linguistics/Ieithyddiaeth Geltaidd. Readings in the Brythonic Languages: Festschrift for T Arwyn Watkins. Amsterdam: John Benjamins; 1990. p. 109-25.

Beach E, Burnham D, Kitamura C. Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. International Journal of Bilingualism 2001;5:221–235.

Béland R, Kolinsky R. One sound heard as two: The perception of affricates in Quebec French by Belgian French speakers. Clinical Linguistics & Phonetics 2005;3:110–117.

Best, CT. Discovering messages in the medium: Speech perception and the prelinguistic infant. In: Fitzgerald, HE.; Lester, B.; Yogman, M., editors. Theory and research in behavioral pediatrics. New York, NY: Plenum Press; 1984.

Best, CT. Learning to perceive the sound pattern of English. In: Rovee-Collier, C.; Lipsitt, L., editors. Advances in Infancy Research. Vol. 9. Norwood NJ: Ablex Publishing Corporation; 1994. p. 217-304.

Best, CT. A direct realist perspective on cross-language speech perception. In: Strange, W., editor. Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research. Timonium MD: York Press; 1995. p. 167-200.

Best CT, Avery RA. Left hemisphere advantage for click consonants is determined by linguistic significance. Psychological Science 1999;10:65–69.

Best CT, McRoberts GW. Infant perception of nonnative consonant contrasts that adults assimilate in different ways. Language & Speech 2003;46:183–216. [PubMed: 14748444]

Best CT, McRoberts GW, Goodell E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. Journal of the Acoustical Society of America 2001;109:775–794. [PubMed: 11248981]

Best CT, McRoberts GW, Sithole N. Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. Journal of Experimental Psychology: Human Perception & Performance 1988;14:45–60.

Browman C, Goldstein L. Towards an articulatory phonology. Phonology Yearbook 1986;3:219–252.

Browman C, Goldstein L. Articulatory gestures as phonological units. Phonology 1989;6:201–251.

Browman C, Goldstein L. Articulatory phonology: An overview. Phonetica 1992;49:155–180. [PubMed: 1488456]

Brown, C. The feature geometry of lateral approximants and lateral fricatives. In: van der Hulst, H.; van de Weijer, J., editors. Leiden in Last: HIL Phonology Papers I. The Hague: Holland Academic Graphics; 1995. p. 41-88.

Calderón J, Best CT. Perceptual assimilation of non-native vowel contrasts to the American English vowel system. Journal of the Acoustical Society of America 1996;99:2602.

Chomsky, N.; Halle, M. The Sound Pattern of English. New York: Harper and Row; 1968.

Clements GN. Phonological primes: Features or gestures? Phonetica 1992;49:181–93.

Clements, GN. Affricates as non-contoured stops. In: Fujimura, O.; Joseph, B.; Palek, B., editors. Proceedings of LP '98: Item order in language and speech. Prague: The Karolinum Press; 1999. p. 271-299.

Clements GN. Feature economy in sound systems. Phonology 2003;20:287–333.

Clements, GN. The role of features in speech sound inventories. In: Raimy, E.; Cairns, C., editors. Contemporary Views on Architecture and Representations in Phonological Theory. Cambridge, MA: MIT Press; 2009. p. 19-68.

Clements, GN. Feature organization. In: Brown, K., editor. Encyclopedia of language and linguistics. 2nd. Vol. 4. Oxford: Elsevier Limited; 2006. p. 433-441.

Clements, GN.; Hume, EV. The internal organization of speech sounds. In: Goldsmith, JA., editor. The Handbook of Phonological Theory. Oxford: Blackwell; 1995. p. 245-306.

Dupoux E, Pallier C, Kakehi Y, Mehler J. New evidence for prelexical phonological processing in word recognition. Language and Cognitive Processes 2001;16:491–505.

Ewen, CJ.; van der Hulst, H. The phonological structure of words. Cambridge UK: Cambridge University Press; 2001.

Fowler CA. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics 1986;14:3–28.

Fowler CA. Real objects of speech perception: A commentary on Diehl and Kluender. Ecological Psychology 1989;1:145–160.

Fowler C, Brown J, Sabadini L, Weihing J. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. Journal of Memory and Language 2003;49:396–413.

Galantucci B, Fowler C, Goldstein L. Perceptuo-motor compatibility effects in speech. Attention, Perception, & Psychophysics 2009;71:1138–1149.

Goldstein L, Browman C. Representation of voicing contrasts using articulatory gestures. Journal of Phonetics 1986;14:339–342.

Goldstein, L.; Byrd, D.; Saltzman, E. The role of vocal tract gestural action units in understanding the evolution of phonology. In: Arbib, M., editor. From action to language: The mirror neuron system. Cambridge: Cambridge University Press; 2006. p. 215-249.

Goldstein, L.; Chitoran, I.; Selkirk, E. Proceedings of the XVIth International Congress of Phonetic Sciences. Saarbrücken, Germany: 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber; p. 241-244.

Goldstein, L.; Fowler, C. Articulatory phonology: A phonology for public language use. In: Schiller, N.; Meyer, A., editors. Phonetics and phonology in language comprehension. Berlin: Mouton-de Gruyter; 2003. p. 159-208.

Halle M, Stevens KN. A note on laryngeal features. MIT Quarterly Progress Report 1971;101:198–212.

Halle M. Feature geometry and feature spreading. Linguistic Inquiry 1995;26:1–46.

Halle M, Vaux B, Wolfe A. On feature spreading and the representation of place of articulation. Linguistic Inquiry 2000;31:387–444.

Hallé PA, Best CT. Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. Journal of the Acoustical Society of America 2007;121:2899–2914. [PubMed: 17550188]

Hallé P, Dominguez A, Cuetos F, Segui J. Phonological mediation in visual masked priming: Evidence from phonotactic repair. Journal of Experimental Psychology: Human Perception and Performance 2008;34:177–192. [PubMed: 18248147]

Hallé P, Segui J, Frauenfelder U, Meunier C. Processing of illegal consonant clusters: A case of perceptual assimilation? Journal of Experimental Psycholology: Human Perception & Performance 1998;24:592–608.

Holton G. Fortis and lenis fricatives in Tanacross Athapaskan. International Journal of Applied Linguistics 2001;67:396–414.

Keating P. Phonetic and phonological representation of stop consonant voicing. Language 1984;60:286–319.

Keating P. A survey of phonological features. UCLA Working Papers in Phonetics 1988;66:124–150. Distributed by the Indiana University Linguistics Club, Bloomington, Indiana.

Keating P. Phonetic representations in a generative grammar. Journal of Phonetics 1990;18:321–334.

Kehrein, W. Phonological representation and phonetic phasing: Affricates and laryngeals. Tübingen: Max Niemeyer; 2002.

Keyser, SJ.; Stevens, KN. Enhancement revisited. In: Kenstowicz, M., editor. Ken Hale: a Life in Language. Cambridge, MA: MIT Press; 2001. p. 271-291.

Kuhl, PK.; Iverson, P. Linguistic experience and the “Perceptual Magnet Effect”. In: Strange, W., editor. Speech perception and linguistic experience: Theoretical and methodological issues. York Press; Timonium, MD: 1995. p. 121-154.

Ladefoged, P.; Maddieson, I. The sounds of the world's languages. Oxford: Blackwell; 1996.

Lahiri A, Reetz H. Distinctive features: Phonological underspecification in processing. Journal of Phonetics. this issue.

Maddieson, I. Patterns of sounds. Cambridge: Cambridge University Press; 1984.

Maddieson I, Smith C, Bessell N. Aspects of phonetics of Tlingit. Anthropological Linguistics 2001;43:135–176.

Mitterer H, Ernestus M. The link between speech perception and production is phonological and abstract: Evidence from the shadowing tasks. Cognition 2008;109:168–173. [PubMed: 18805522]

Nam, H.; Goldstein, LG.; Saltzman, E. Self-organization of syllable structure: A Coupled oscillator model. In: Pellegrino, T.; Marisco, E.; Chitoran, I., editors. Approaches to phonological complexity. Berlin: Mouton-de Gruyter; 2009.

Raphael, L.; Tobin, Y.; Faber, A.; Most, T.; Kollia, H.; Milstein, D. Intermediate values of voice onset time. In: Bell-Berti, F.; Raphael, L., editors. Producing Speech: Contemporary Issues for Katherine Harris. AIP Press; 1995. p. 117-127.

Sagey, E. Doctoral dissertation. MIT; Cambridge, MA: 1986. The representation of features and relations in non-linear phonology.

Sharma A, Marsh CM, Dorman MF. Relationship between N1 evoked potential morphology and the perception of voicing. The Journal of the Acoustical Society of America 2000;108:3030–3035. [PubMed: 11144595]

Shin SJ. Cross-language speech perception in adults: Discrimination of Korean voiceless stops by english speakers. Studies in the Linguistic Sciences 2001;31:155–166.

Steriade, D. Closure, release, and nasal contours. In: Huffman, MK.; Krakow, RA., editors. Phonetics and Phonology vol 5: Nasals, Nasalization, and the Velum. New York: Academic Press; 1993. p. 401-470.

Stevens, K. Phonetic features and lexical access. In: Fujisaki, H., editor. Recent research towards advanced man-machine interface. New York: Elsevier; 1996. p. 267-280.

Stevens K. Toward a model for lexical access based on acoustic landmarks and distinctive features. Journal of the Acoustical Society of America 2002;111:1872–91. [PubMed: 12002871]

Studdert-Kennedy, M.; Goldstein, L. Launching language: The gestural origin of discrete infinity. In: Christiansen, M.; Kirby, S., editors. Language evolution: The states of the art. Oxford UK: Oxford University Press; 2003. p. 235-254.

Tees RC, Werker JF. Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. Canadian Journal of Psychology 1984;38:579–590. [PubMed: 6518419]

Vaux B. The laryngeal specifications of fricatives. Linguistic Inquiry 1998;29:497–511.

Werker JF, Gilbert JHV, Humphrey GK, Tees RC. Developmental aspects of cross-language speech perception. Child Development 1981;52:349–355. [PubMed: 7238150]

Williams L. The voicing contrast in Spanish. Journal of Phonetics 1977;5:169–184.

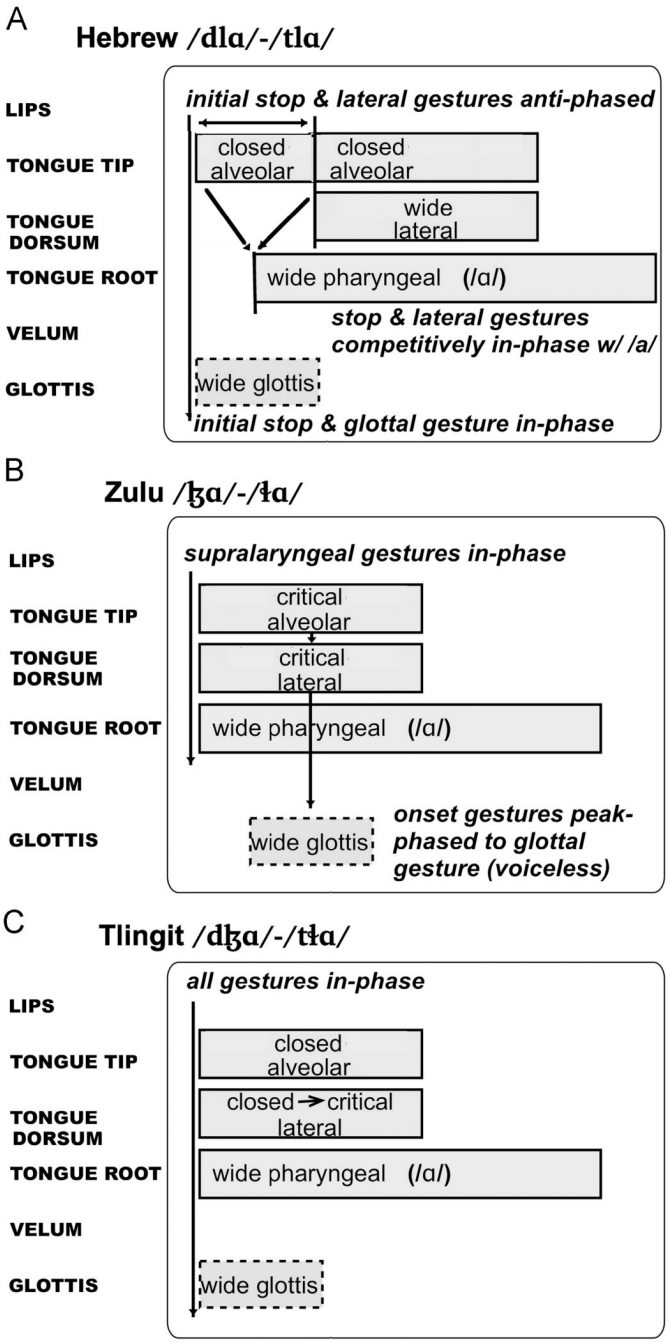


Figure 1.

Schematic diagrams of posited gestural phasing relations and complexity for the three target voicing contrasts: (A) in Hebrew /dl/-/tl/ the glottal gesture for the voiceless stop is in-phase with the initial coronal stop gesture, but the stop gestures are anti-phased with the lateral gesture, and both the stop and lateral gestures compete to be in-phase with the vowel (/a/) which shifts the phasing leftward from 0° for the stop and rightward from 0° for the lateral; (B) in Zulu /ɮɑ/-/ɬɑ/ the lateral fricative gestures are all in-phase, but they are peak-phased with the glottal gesture for the voiceless case; C) in Tlingit /dɮɑ/-/tɬɑ/ all gestures including the glottal gesture are in-phase (onset- rather than peak-phased).

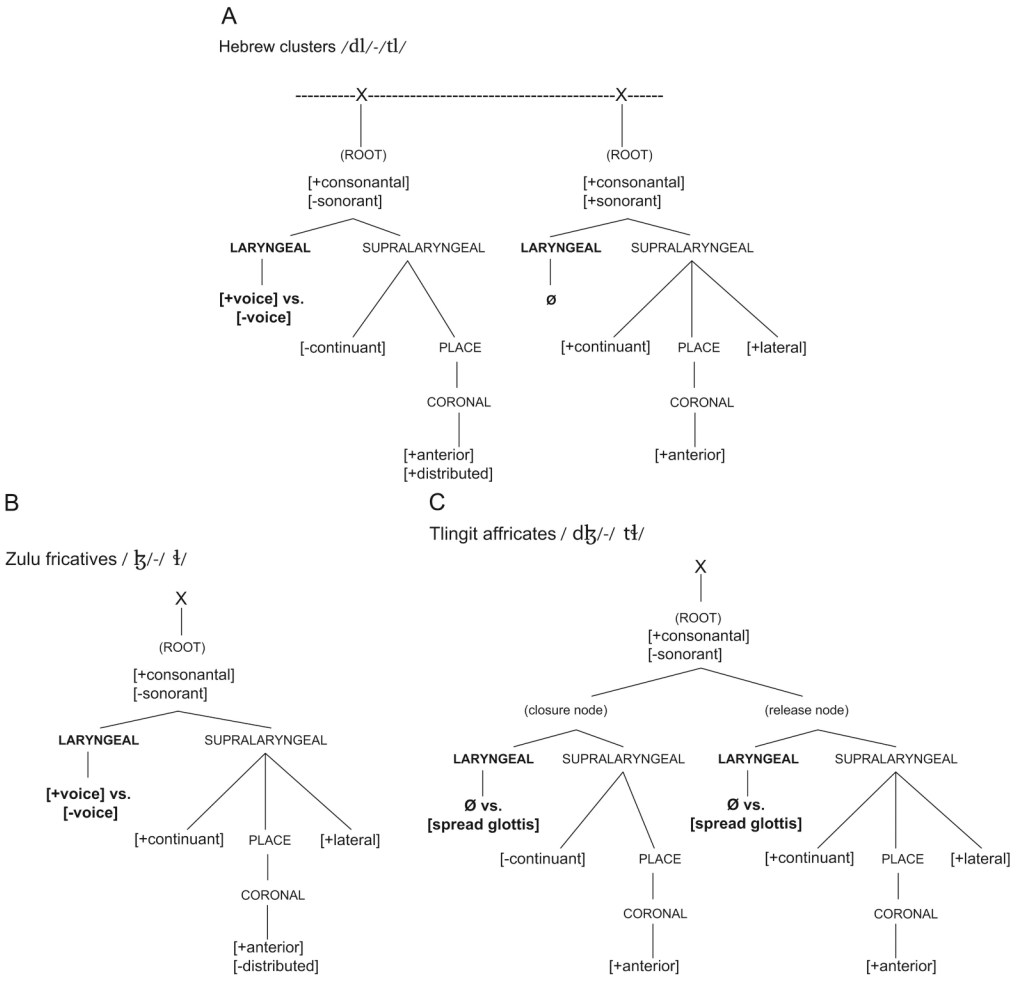


Figure 2.

Schematic diagram of posited feature tree structures for the three target voicing contrasts: (A) Hebrew /dl/-/tl/; (B) Zulu /ɮa/-/ɬa/; (C) Tlingit /dɮa/-/tɬa/. A single tree structure is shown for each language, with the contrastive features for the voicing distinction shown in boldface.

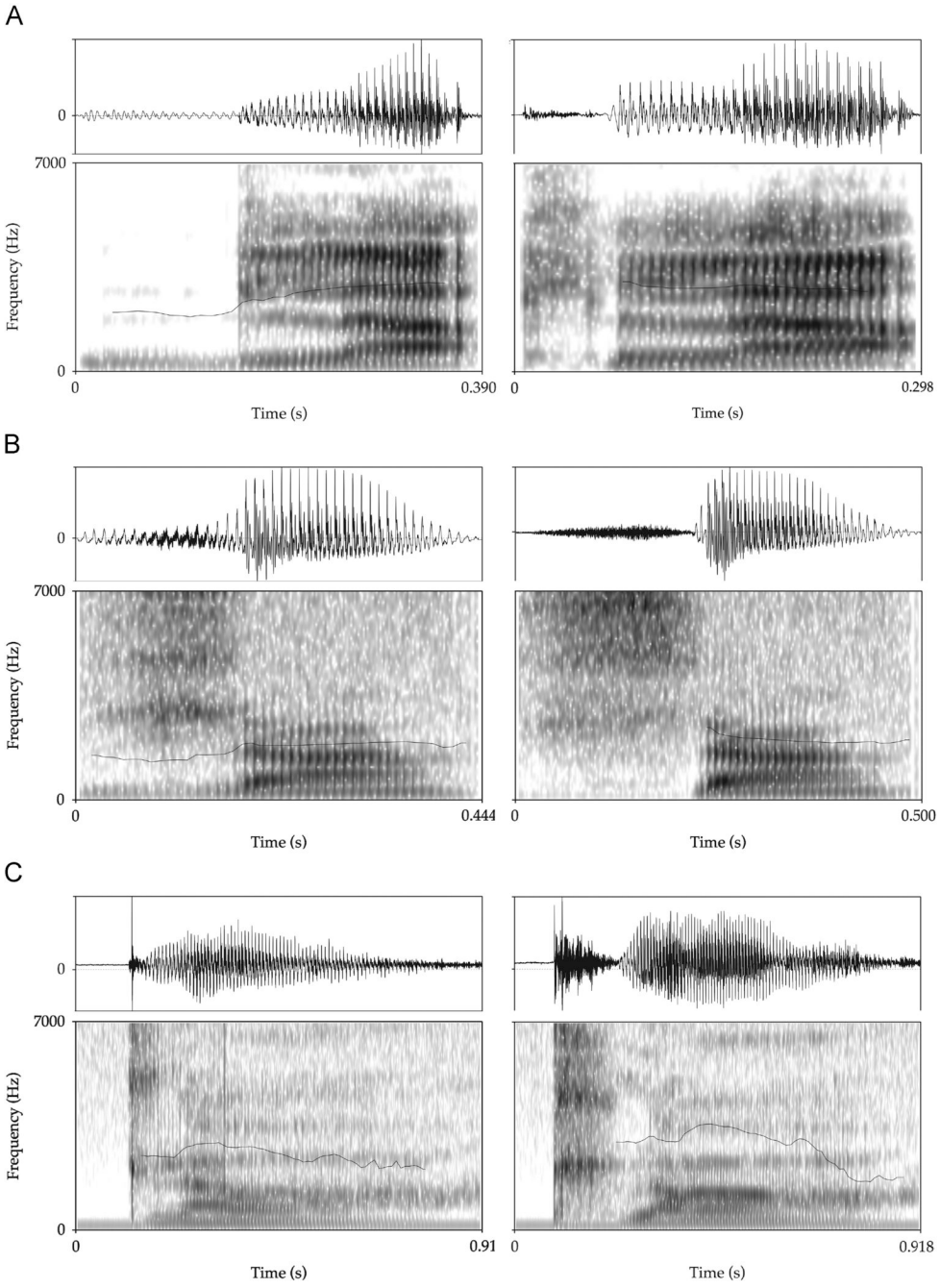


Figure 3.
Example waveforms (upper panel in each row) and spectrograms (lower panel in each row) for tokens of each of the six target onset types: (A) Hebrew /dl/-/tl/; (B) Zulu /ɮa/-/ɬa/; (C) Tlingit /dɮa/-/tɬa/ (left column = phonologically voiced; right column = phonologically voiceless).

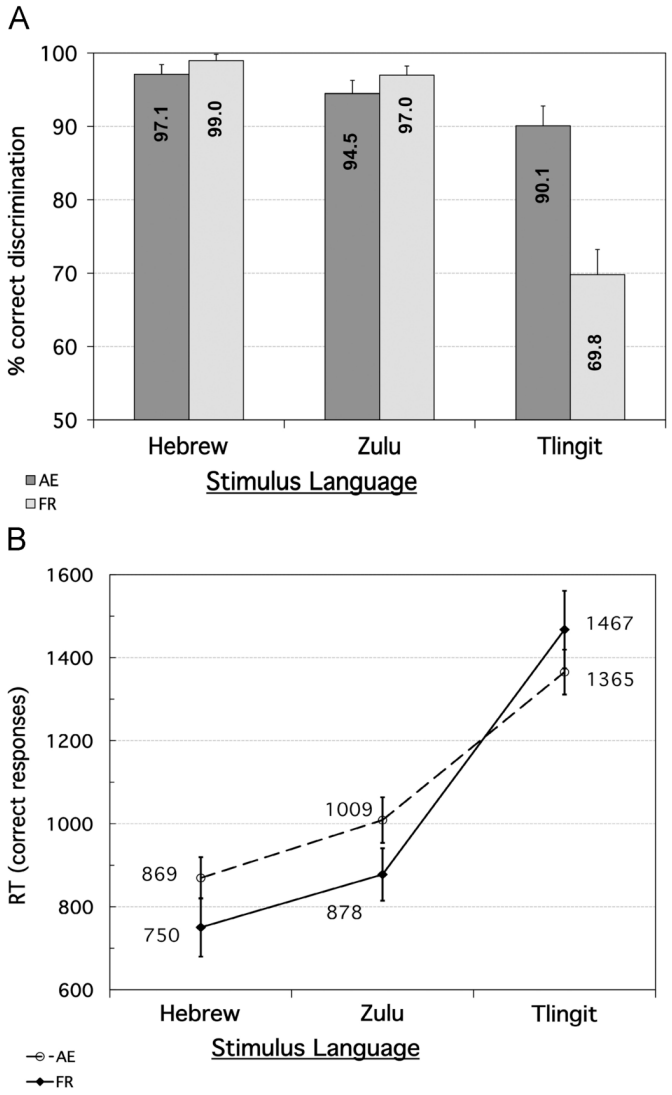


Figure 4.

Results for English (AE, American English) and French (FR, Parisian French) listeners in the AXB discrimination task: (A) mean percent correct discrimination (standard error bars); and (B) mean reaction time (RT) in ms for correct responses (standard error bars).

Table 1

Schematic overview of gestural complexity and tightness of intergestural phasing for the three types of onset contrasts, from an articulatory perspective (see Figure 1).

gestural complexity:	Hebrew > Tlingit > Zulu
tightness of intergestural phasing:	Tlingit > Zulu > Hebrew