ANNUAL REVIEWS

Annual Review of Linguistics

How Consonants and Vowels Shape Spoken-Language Recognition

Thierry Nazzi[1] and Anne Cutler[2,3]

[1]Laboratoire Psychologie de la Perception, CNRS–Université Paris Descartes, 75270 Paris Cedex 06, France; email: thierry.nazzi@parisdescartes.fr

[2]The MARCS Institute for Brain, Behaviour and Development and ARC Centre of Excellence in the Dynamics of Language, Western Sydney University, Penrith, New South Wales 2751, Australia; email: A.Cutler@westernsydney.edu.au

[3]Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands

Annu. Rev. Linguist. 2019. 5:9.1–9.23

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-011718-011919

Keywords

consonants, vowels, phonology, lexicon, speech recognition, adult, infant, development

Abstract

All languages instantiate a consonant/vowel contrast. This contrast has processing consequences at different levels of spoken-language recognition throughout the lifespan. In adulthood, lexical processing is more strongly associated with consonant than with vowel processing; this has been demonstrated across 13 languages from seven language families and in a variety of auditory lexical-level tasks (deciding whether a spoken input is a word, spotting a real word embedded in a minimal context, reconstructing a word minimally altered into a pseudoword, learning new words or the "words" of a made-up language), as well as in written-word tasks involving phonological processing. In infancy, a consonant advantage in word learning and recognition is found to emerge during development in some languages, though possibly not in others, revealing that the stronger lexicon–consonant association found in adulthood is learned. Current research is evaluating the relative contribution of the early acquisition of the acoustic/phonetic and lexical properties of the native language in the emergence of this association.

Review in Advance first posted on August 1, 2018. (Changes may still occur before final publication.)

1. INTRODUCTION

The contrast between vowels and consonants[1] has a special status in the study of language. It is a true linguistic universal in that no natural language lacks either category (Ladefoged & Disner 2012). It is not an easy contrast to define, because speech sounds can be conceived of as falling along a sonority hierarchy, from monophthongal vowels as the most sonorous to brief transitional stop consonants as the least. Nonetheless, all languages contain phonemes at distant positions on that continuum, which entails that all languages end up with a vowel/consonant contrast, bringing processing consequences. In this review, we trace the story from the phonetic structure to its inevitable perceptual implications, and the associated developmental challenges.

The research we survey is, as the title indicates, entirely perceptual. But the V/C contrast itself is one of speech production: The flow of air through the vocal tract is unimpeded in the case of V, but partially or fully obstructed for a C. Acoustically, a V is thus continuous and can last as long as a breath can last, while a C is transitional and hence transitory. Mouth movements (of infants or adults) accompanied by phonation result in a typically cyclic pattern with transitional states interrupting periods of continuity.

Language phonology interprets this alternation as a sequence of speech sounds. The selection of the sound categories, their grouping into syllables, and the constraints on their sequencing are all language specific. Although it is an articulatory fact that a speaker can in effect utter one V and seamlessly progress into another, and many language repertoires contain diphthongs as well as monophthongs (Maddieson 1984), not all do. Likewise, because C articulation interrupts air flow in different ways and at different positions in the vocal tract, it is less easy to articulate C jointly or sequentially.[2] Thus, some languages apply sonority sequence constraints to C sequences that can function as a single unit (these are mostly affricates; [tʃ] as in *ditch*, is an affricate, but [ʃt], as in *dished*, is not), and to C sequences in syllable onsets or codas. But other languages bar within-syllable C sequences entirely (such that, for instance, loan words from less fastidious languages may be reanalyzed as containing transitional vowels; English *express* becomes Japanese *ekusupuresu)*. In this review, we describe how further consequences of the articulatory distinctions constrain both adult speech perception (Section 2) and acquisition of spoken language (Section 3), and also how the language-specific choices of particular phonologies further shape processing decisions and indeed the architecture of the speech processing system.

2. ADULT PROCESSING

2.1. Vowel and Consonant Perceptibility

For perceivers, the articulatory differences create effective categoricality differences. These differences are observed in phonetic categorization studies, in which an artificial continuum is constructed between two clear phoneme endpoints, and listeners attempt to discriminate tokens along this continuum or judge whether tokens better match one or the other endpoint. Stop consonants tend to be judged categorically; first, listeners readily discriminate stimuli drawn from different stop categories but fail to discriminate stimuli drawn from the same category, even when acoustic differences between stimuli are comparable, and second, stop identification delivers a typical

[1]Because this whole review concerns vowels and consonants, and readers might rapidly tire of reading the same words so often, we often use V for vowel(s) and C for consonant(s).

[2]The wide variety of language realizations includes some phoneme repertoires that defy this generalization, such as the double articulations observed in the Pacific language Yélî Dnye. Alas, little speech perception evidence exists for this or similarly endowed languages.

"categorical" function in which decisions favor the nearest endpoint category except around the continuum midpoint, where the function shifts steeply between categories. Vowel continua, by contrast, show neither pattern; identification displays a more gradual category shift along the whole continuum, and discrimination is more sensitive even within categories (Pisoni 1973).

Likewise, the articulatory differences influence the phonetic realization of spoken targets (and, in turn, relative perceptibility). In V production, the vocal tract adopts a particular target shape; but closures for preceding and following C affect these V configurations, which, although in total longer than C, may hardly attain their target shapes at all. This means that, effectively, V articulation is affected by C articulation to a greater degree than vice versa. Perceptual effects of this are described in Section 2.3. By contrast, greater clarity of V signals (Repp 1984) delivers perceptual advantages such as lower V rates in slips of the ear (Bond & Garnes 1980), better V resistance to noise masking [Nooteboom & Doodeman 1980; note that listeners rely on different cues for perceiving V and C in noise (Parikh & Loizou 2005)], and no ear asymmetry for V perception but a right-ear advantage when C is added, that is, for CV syllables (Shankweiler & Studdert-Kennedy 1967). If phonemically balanced sentences are presented with only their V being audible versus only obstruent C (Cole et al. 1996) or any C (Kewley-Port et al. 2007), with or without pitch information (Fogerty & Humes 2012), the V-preserved versions always prove easier to recognize. A speech production consequence is that speakers tend to talk more loudly in languages with fewer V (thus, with higher functional load for C; Nettle 1994).

In larger contexts, a sentence or more, these effects result in a V-advantage over C, at least with respect to resistance to error. However, in more restricted contexts (a word, a syllable), V/C differences are not always found. Words spoken after a prime differing in one phoneme are recognized equivalently whether the difference is in a V or a C (Cutler et al. 1999), and the same holds true for primes that are only fragments of the target word (Soto-Faraco et al. 2001). Performing phoneme tasks on C is harder if V varies (e.g., /b/–/d/ discrimination in *ba du de bi* versus *ba da da ba*, or /d/ detection in *ba ge di fo* versus *ba ga da fa*), and the same situation holds for the inverse, that is, discriminating and detecting V targets with varying versus constant C (Wood & Day 1975, Swinney & Prather 1980).

But for lexical encoding, C shows an advantage. Creel et al. (2006) taught listeners pseudowords paired with unfamiliar pictures, and then asked them to choose a picture's correct name (e.g., *pibo*) among foils (*pibu*, *pabu*, *dibo*, *diko*). C-preserved items proved more confusable than V-preserved items, again pointing to C frames as the strongest parts of the newly learned names. Havy et al. (2014) and Escudero et al. (2016) also found more accurate encoding of C than of V in learning of pseudoword minimal pairs.

Interestingly, when there is asymmetry, it can vary across languages. Phoneme detection is faster and more accurate for C than V in English (van Ooijen et al. 1991), but not in Japanese (Cutler & Otake 1994). In Dutch (with similarly large C and V inventories), variation interference on categorization is symmetrical in size for C and V responses, but in Spanish (with four times as many C as V), V responses are more affected by varying C than vice versa (Costa et al. 1998). The latter finding shows not only that perception is affected by phoneme repertoire distribution, just as Nettle (1994) showed production to be, but also that listening processes adjust to the native-language potential for relative contextual variability.

The most common V repertoire size across languages is five (Maddieson 1984); most repertoires contain many more C than V, making C distinguish words better than V. In French, for instance, knowing the identity and order of the consonants of a CVCVCV word (e.g., [k_s_n in *casino*) restricts the average number of possible candidates to 6.03 words, while knowing the vowel sequence (e.g., [a_i_o]) produces an average of 8.8 possible candidates: 1.46 times as many (Delle Luche et al. 2014, Keidel et al. 2007). Syllable structure constraints exacerbate the effect;

permitted sequences are longer for C than V, so words become more likely to contain more C than V. Moreover, lexical V harmony processes, across languages, are more common than C harmony, and morphological processes also affect V more than C, all making C, again, the more reliable information. In Semitic languages, word families can share a consonant framework with the vowels distinguishing the family members (e.g., *book*, *writing*, and *scribe* share the consonants [ktb] in Arabic); no language features the reverse mapping.

Analyzing the structure of a vocabulary could thus reveal the relative informativeness of C and V categories in the lexicon. Although the framework-plus-embellishment model works most clearly for Semitic languages, this C/V asymmetry is more generally true, is machine-learnable (Monaghan & Shillcock 2003), and, as the following sections document, has consequences for spoken-language recognition.

2.2. Reconstructing Altered Words

Strong support for the notion that C and V play different roles in defining lexical items comes from a task termed word reconstruction (WR), developed by van Ooijen (1996). In a WR experiment, listeners hear pseudowords such as *shevel* [ʃɛvəl] and attempt to transform them into a real word by altering a single sound. The stimuli can become words with change in a single C or V—so *shevel* could become either *level* or *shovel*. Van Ooijen discovered that listeners more often opted for a V change, and their response times were faster for V than for C changes. When instructed specifically to change a C, they made more errors and responded more slowly than if a V alteration was requested. The results were not affected by the relative number of V and C in each pseudoword (*shevel*, with more C than V; *ezzy* [ɛzi] *easy/any* with more V than C; *eluff* [əlʌf] *aloof/enough* with the same number of each).

This picture is clearly compatible with an explanation in which listeners treat the pseudowords as two frameworks, with the consonantal frame constraining lexical identity more than the vocalic frame. Van Ooijen, noting the asymmetry in phonetic context effects for C and V described in Section 2.1, concluded that listeners treat V as more mutable objects than C.

Van Ooijen's results for British English have been replicated in Dutch (Cutler et al. 2000), in Japanese (Cutler & Otake 2002), and in both Castilian and American varieties of Spanish (Cutler et al. 2000, Marks et al. 2002). These replications ruled out several classes of alternative explanations, invoking phoneme repertoire makeup (selection of a candidate phoneme for alteration might seem easier from a smaller as opposed to a larger set, but Dutch has very comparable V and C repertoire sizes), dependence upon a particular phonology (Spanish has no vowel reduction; Japanese has no lexical stress; English has both), or dialect experience (in contrast to English dialects, Spanish dialects vary more in C than in V). A further replication in American English (Moates et al. 2002) added the demonstration that token frequency of C and V segments was unrelated to outcome patterns. Finally, a brain imaging study in British English (Sharp et al. 2005) showed that WR involving either C or V substitution engaged the same known lexical processing areas in the brain,[3] but C substitution, being again more difficult than V substitution, caused greater cortical activation.

The sole failure to observe a C/V asymmetry in WR is a study in Mandarin in which all stimuli were monosyllabic pseudowords such as *su3* and *jiong4* (Wiener & Turnbull 2016). In this

[3] Caramazza et al. (2000) report a double dissociation between two patients with aphasia, involving, respectively, the disruption of C production with unimpaired V production and the reverse pattern. Note that neither patient showed related effects in speech perception.

study, accuracy and response time were compared for changes in the syllable's tone, C, or final.[4] Altering the tone turned out to be by far the easiest response, in line with many earlier findings in a wide range of tasks that tones constrain lexical identity less well than segments (e.g., Tsang & Hoosain 1979, Taft & Chen 1992, Cutler & Chen 1997, Ye & Connine 1999). However, Wiener and Turnbull observed no significant difference between the conditions in which V(final) versus C alterations were required. This might be an ordering effect (as C always preceded V in the stimulus, and hence offered the earliest option for change; note that order has generally been controlled in WR studies). But it could also, interestingly, indicate the limits on C-advantages in constraining lexical structure: One C doth not a framework make. If a C/V asymmetry were to re-emerge in further studies with other Mandarin stimuli such as polysyllabic lexical items, this could help illuminate exactly how consonant frames guide lexical access.

2.3. Segmenting Artificial Languages into "Words"

The C-advantage in signaling lexical identity appears in artificial language learning (ALL). Listeners hear a continuous string of syllables made up of a number of fixed syllable sequences, most often CV syllables making a handful of three- or four-syllable "words." The string *dutazidavotuzovidotivuzazovidotivuzadutazidavotudutazitivuzazovido* contains the "words" *davotu*, *tivuza*, *zovido*, and *dutazi*, made of just four C and four V (see, e.g., Saffran et al. 1996).

The transitional probability of the syllables in such a string enables listeners to group the recurring sequences, and hence learn the set of words. Peña et al. (2002) examined nonadjacent syllable-level dependencies (embodying segment-level constraints: Frame elements began with stops; variable elements began with continuants). "Words" were learned, but underlying dependency rules were not, unless additional segmentation cues such as subliminal interword silences (25 ms) were added; event-related potential (ERP) findings by Mueller et al. (2008) supported this observation. Newport & Aslin (2004) showed that dependency could be learned at the phonemic level, as actually occurs in many phonologies—either Semitic-style C frames (p_g_t, d_k_b) with variable V, or V sets (a_e_u, i_o_a), with variable C, mimicking vowel dependencies found in natural languages. Without Peña et al.'s segmental constraints, however, no syllable-level learning appeared; so Newport and Aslin concluded that segmental-level regularities are what listeners find most natural to deal with. Further explorations (Perruchet et al. 2004, 2006; Onnis et al. 2004, 2005) confirmed the priority of segmental regularities, and eventually prompted general agreement that C frames, particularly with initial stops, constitute the strongest regularity for word-level units (Bonatti et al. 2005, Onnis et al. 2005, Keidel et al. 2007, Toro et al. 2008). Mehler et al. (2006) constructed clever ALL materials containing equally probable C and V frames; the former proved significantly easier to learn.

Although agreement was reached on the importance of C for lexical processing, the protagonists in this debate could not concur on the origin of this C-advantage: learned lexical statistics prompting language users to treat C and V differently (Seidenberg et al. 2002, Keidel et al. 2007); or, C and V being treated differently in perception and phonological processing, with the lexical statistics following from that asymmetry (Bonatti et al. 2005, 2007; Mehler et al. 2006; Onnis et al.

[4]Wiener & Turnbull's (2016) task was altering not C and V but rather the initial versus final part of the syllable, conforming to Chinese phonology and preferred sound structure descriptions. Of the 96 syllables tested, 24 were CV (*su*), 40 were CVV (*kua*), and 32 had a final consisting of V(V) plus nasal coda (*sun*, *liang*). Substitution of a final may not match in complexity to substitution of a single vowel in European languages; Wiener and Turnbull do not report results separately by test-syllable final type, nor do they separate response types (e.g., *le* becoming *la*, *lei*, or *liang*, or vice versa).

2005). No ALL results forced either conclusion. As argued in Section 3, below, however, evidence from spoken-language acquisition may prove more informative.

Also strikingly, agreement concerned primarily the importance of C (for processing words). A matching role for V was never as clear. Proposals for the role of V have invoked grammar (Nespor et al. 2003, Bonatti et al. 2005) or prosody (Mehler et al. 2006, Kolinsky et al. 2009), but relevant evidence of V processing advantages for these purposes seems harder to capture. Some other types of evidence involving tasks of the kind described in Section 2.1 may help, however.

Thus, Kolinsky et al. (2009) manipulated pitch pattern and phonemic structure in a variation interference study. Same–different judgments of C structure were independent of pitch variation, but judgments of V structure were not, consistent with the role of V as carrier of prosody. Same–different judgments for pairs of words in a study by Owren & Cardillo (2006) concerned either the semantic relatedness of the words or the speaker who spoke the words; the stimuli were distorted in that only C or V was audible (see, e.g., Kewley-Port et al. 2007). Accuracy was higher for speaker decisions on V stimuli (and for semantic decisions on C stimuli), consistent with a role for V in speaker recognition (note that speaker variation is particularly expressed in V; Johnson et al. 1993). Section 3 includes converging evidence for this result.

2.4. Segmenting Real Words from Minimal Contexts

Evidence thus suggests an advantage for V in longer communications. While minimal communications can consist solely of a C (e.g., *shhh!* as a suggestion to be quiet), a general rule is that utterances must minimally be a syllable. Across languages, syllables consist of a nucleus plus preceding (onset) and following (coda) C material. Languages constrain syllable structure differently (in what may function as a nucleus, in how complex onsets or codas may be, in whether codas are allowed at all). However, the minimal syllable is a nucleus, and by far the most common case is that syllabic nuclei must be vocalic. Thus, a V is more likely than a C to be required in every utterance.

Listeners employ such a probability calculation in how they find the words of which spoken utterances are composed, as shown by an experimental task called word-spotting, designed to focus on the process of segmenting continuous speech into its component units. In the nonsense string *obzel sleepnah glershath chimdowb maffegg hezzing lumpev*, participants would be expected to spot *sleep*, *egg*, and *lump*. Spotting words is easier if they occur in a syllabic, and hence V-bearing, context such as these, as opposed to a context with no vowel (e.g., *sleept*, *fegg*, or *blump*; Norris et al. 1997). This does not mean that speech segmentation is a syllabic procedure, because if the task is syllable detection, no such context effect appears (the syllables *heap* or *feep* are detected equally easily in syllabic or consonantal contexts such as *heapnah* or *feept*; Kearns et al. 2002). The +V/−V difference is specific to word segmentation.

The difficulty with V-less contexts first discovered in British English is quite general: It appears also in American English (Newman et al. 2011), Dutch (McQueen & Cutler 1998), German (Hanuliková et al. 2011), French (Dumay et al. 2002, using a slightly different task), Japanese (McQueen et al. 2001), and Cantonese (Yip 2004a,b; again with an adapted task). It could also be determined that it was the presence of a V, rather than syllabic legality, which enables contexts to support word segmentation. The simple presence of a V does not always create a viable syllable. In English, CV sequences where the V is lax (and not schwa) are not syllables; *get* and *stub* are words, but coda-less *ge* and *stu* with the same vowels are prohibited. Yet such syllable legality is irrelevant: Both *ladeserve* and *l*[ɛ]*deserve* allow detection of *deserve*, although [lɑ] is a legal English syllable but [lɛ] is not (see Norris et al. 2001 for British English and Bishop & Toda 2012 for American English).

Higher-level constraints on whether syllables can serve as stand-alone words are also irrelevant. As noted above, many language-specific constraints on syllable and word structure exist; one such is a bimoraicity constraint on word status (McCarthy & Prince 1995). This will obviously turn into a bisyllabicity constraint if syllables can only have one mora; in other words, long V and codas are excluded. That is the case in some Bantu languages, such as Sesotho, where *rora* (*roar*) can be and is a word, but *ro* and *ra* could not be stand-alone words. In Sesotho word-spotting, though, monosyllables facilitate detection of a target word as effectively as bisyllabic contexts do; *rora* is detected as rapidly in *jirora* as in *hapirora* (Cutler et al. 2002a). Again it is crucial that a V is present, but conformity to language-specific syllable-structure constraints is not relevant.

Often, though, syllables can for one or another reason be V-less. Syllabic sonorants, as in the second syllables of English *awful* or *often*, are the most frequent type of syllabic consonant (and are interchangeable with a schwa + C pronunciation). Vowels can be deleted in such a way that the syllabic structure of a word effectively changes—English *family* pronounced *fam'ly*, for instance, or French *galerie* becoming *gal'rie*. More importantly for speech segmentation, V can be deleted such that a consonant-only syllable results; Japanese vowel devoicing is one such case (in *sukiyaki* the first vowel is devoiced; note that *ski* is not a legal syllable in Japanese). This indeed slows word-spotting (Cutler et al. 2009). Even more challenging is the case of wholly vowel-less noncontinuant words, such as function words in some Slavic languages: The Slovak preposition translated as English *to*, for example, is the plosive *k*. Interestingly, word-spotting experiments in Slovak (Hanuliková et al. 2010) revealed that consonants other than those used as function words were difficult contexts, so, *palec* 'finger' was harder to spot when preceded by a pseudoword consonant, *špalec*, than by a whole syllable, *lafpalec*, but a preceding function-word consonant, obstruent, or plosive, even if always meaningless in combination with the target word—*fpalec* 'in finger'—had no such adverse word-spotting effect.[5]

Most challenging of all for this account of the role of C and V in word segmentation are languages in which any sound in the phoneme inventory can function as a syllabic nucleus. The Berber group of languages in North Africa allows this (Ridouane 2008). And two Berber languages, Tashelhiyt and Tarafit, proved to be the first cases where the word-spotting disadvantage for C contexts was found not to appear at all. Detection of, say, *fad* 'thirst' in *ghfad* versus *aghfad* did not significantly differ in accuracy or response time (El Aissati et al. 2012).

The ubiquity and consistency of these findings across languages suggest that the V/C distinction has a role to play in everyday listening, in the process of extracting word forms from running speech. Lexical statistics demonstrate that it would be an efficient strategy for English, as it would exclude from consideration most fully embedded words such as *egg* in *leg* or *two* in *tomb*. Of all the embedded words occurring in the English lexicon, 73% leave a V-less residue in this manner (Cutler et al. 2002b). Use of this strategy in listening to English speech would thus automatically rule out a significant proportion of potentially competing words, and consequently speed up listening. Only where the rejection process would be inefficient, in that it would rule out actual words (e.g., in Berber languages), is it not used.

[5]For completeness, **Table 1** includes a word-spotting study (Alexeeva et al. 2017) in Russian, which has function words like those of Slovak. As in Slovak, function-word consonants did not here slow target-word detection. However, recognition was also not easier with syllable contexts. All target words in this study had iambic stress. With preceding syllable contexts, the most likely stimulus pronunciation pattern is SWS, making the context prosodically stronger than the target word's initial syllable, and the word thus harder to detect in the syllable context condition. In word-spotting studies where target items are recorded as a whole (context + word), rather than contexts being spliced onto separate recordings of the words, it is important to ensure that target words are equally recognizable across contexts. Usually, words are removed from their contexts and these truncated forms are presented for lexical decision. This was done in all such studies listed in **Table 1** except, alas, this Russian study, which thus needs to be repeated in an improved version.

2.5. The Red Herring of Reading

Words are words whether spoken or written; but in reading (conjunct availability) and listening (sequential availability), implications of C/V information load differ. It is beyond the scope of our review to explore this topic further. Meanwhile, findings of reading research on C/V are mixed.

Some effects translate across modality. Word reconstruction with printed materials produces the same results as described for speech; INVATE or HUNDLE becomes INVITE or HANDLE more easily than INVADE or BUNDLE (Moates & Marks 2012), with error rates of both English and Spanish participants fully mimicking the results with spoken materials. In many reading tasks, C letters are read more rapidly than V letters, but possibly because they are visually more distinct (Lee et al. 2002). And, of course, informal evidence (from jokes, from crossword puzzle solvers, from the inventors of texting "abbrvns") suggests that printed C frames outdo V frames in evoking a word; "CNSNNT" allows better recognition of *consonant* than "OOA" does.

Visual priming studies, however, produce conflicting results. In an oft-used method in visual word processing research, a briefly presented prime is replaced almost immediately by a target word on which a decision (word or not a word) must be made. In general, participants are unaware of the prime, but prime–target relationships influence decisions about the target. With this method, Carreiras et al. (2009) confirmed that words are better evoked by preserved C than V letters, and also showed that the C and V prime types induce different brain response patterns. New et al. (2008) and New & Nazzi (2014) found that accepting a visual string of letters (e.g., JOLI 'pretty') as a French word was faster after a C-overlapping written prime (JALU) than after a control prime, but a V-overlapping prime (VOBI) provided no such facilitation. These findings clearly resemble the speech results in which only C frames assist lexical access.

By contrast, in visual lexical decision, nonwords with transposed C letters (e.g., JUGDE) are misidentified as real words more often, and rejected as the nonwords they actually are more slowly, than nonwords with substituted letters (e.g., JUNPE; Andrews 1996). Masked priming typically delivers stronger facilitation if primes contain such a transposition involving two C versus two V (e.g., CONDISER or CINSODER as a prime for CONSIDER; Lupker et al. 2008, Perea & Acha 2009). These findings seem to show, contrary to the priming results, that C frames can be jumbled up with little effect on lexical access, but V sequences cannot. In that case, visual studies would be eliciting a pattern opposite to what is observed in speech experiments.

However, consider that in these transposed-letter reading experiments, the C/V contrast is not usually phonological. A "vowel" is one of the letters A, E, I, O, or U, and everything else is a "consonant." Also, masked transposed-letter primes that alter the pronunciation of the transposed letters do not differ at all in their effect from primes without such influence (as in, e.g., Spanish pairs such as RACIDAL for RADICAL, where the letter *c* would be pronounced differently before *i* versus *a*, versus FRASACO for FRACASO, where no change in letter pronunciation results; Perea & Carreiras 2008). This failure to find an effect of pronunciation indicates that a target word's phonological form plays no role in these experiments.

Furthermore, as pointed out by Schubert et al. (2018), letter transposition typically confounds letter position with the nature of adjacent letters; consonants (but not vowels) abut another consonant in one presented form but not in the other (compare ALCOHOL with its prime ALHOCOL). Ruling out this factor (e.g., with items such as LUNATIC primed by LUTANIC) causes the C/V difference in masked transposed-letter priming to disappear.

We suggest that the visual and spoken studies do not in fact conflict. If a real word is to be consciously sought and phonology helps [as in some lexical decision paradigms (New et al. 2008) or in word reconstruction (Moates & Marks 2012)], then a preserved consonantal frame facilitates. But if the phonology condition is not met (and the task is performed at the orthographic level; see New & Nazzi 2014 for a discussion), no such facilitation occurs. Indeed, although cross-modal priming is a reliable psycholinguistic technique (Zwitserlood 1996), spoken similar pseudoword primes (*cummel* for CAMEL or CUDDLE) do not necessarily affect responses to visually presented target words at all (Cutler et al. 1999); and, as noted above, the most striking case of a reversed C/V asymmetry in the visual mode, transposed-letter priming, vanishes if stimulus structure is controlled (Schubert et al. 2018).

Finally, consider two tasks used by Carreiras & Price (2008) in a brain imaging study with transposed-letter printed stimuli: reading aloud versus lexical decision. The former (speech production) task activated a right-hemisphere area associated with prosodic processing, causing greater activation when stimuli had transposed vowels (e.g., PRIMEVARA for PRIMAVERA 'spring') versus consonants (PRIVAMERA). The lexical decision task, in contrast, activated a right-hemisphere area linked to response inhibition (in a way consistent with the visual but not the spoken findings; items with jumbled C were harder to reject), but did not activate any area associated with lexical processing (such as the lexically important left inferior frontal gyrus active in spoken word reconstruction; Sharp et al. 2005). The most probable explanation of this outcome is that, in line with results obtained by Perea & Carreiras (2008) and the findings listed immediately above, printed-word recognition does not need to engage the mechanisms underlying the strong C/V asymmetries in lexical processing. Consonants and vowels are acoustic entities, not written ones; the consonant advantage for lexical recognition involves listening, not reading.

2.6. Adult Processing: Overview

Table 1 summarizes 38 studies of adult auditory lexical processing reviewed in the preceding sections. The studies involved 13 languages from 7 language families and used a variety of tasks: deciding whether a spoken input is indeed a word, spotting a real word embedded in a minimal context, reconstructing a word that has been minimally altered into a pseudoword, and learning new words or the "words" of a made-up language. These tasks mostly revealed a stronger association for consonants whenever lexical processing was involved. Predicted exceptions were observed when the language's phonology did not constrain C/V roles in syllable structure (El Aissati et al. 2012) and when the task did not involve word decisions (Kearns et al. 2002). Most noticeable is that no study of any kind in any language showed a stronger lexical relation for V alone.

The past quarter-century of research on the roles of C and V in adult spoken-language processing has thus yielded clear support for a differentiation at the lexical level, which, to a considerable degree, follows from the articulatory/acoustic structure of speech. Above the lexical level, V patterning carries prosodic structure and supports processing at levels higher than the word, although here the empirical evidence is scarcer and the link to phonetic structure less clear cut. Also scarce, we point out, is evidence from languages with simple syllabic phonology and/or lexical tone; intriguingly, such evidence as is so far available is consistent with the possibility of different and more complex dependencies (Wiener & Turnbull 2016, Gomez et al. 2018, Poltrock et al. forthcoming). We hope that these gaps in the literature will soon be filled.

In Section 3, we turn to the developmental evidence; given that language-specific lexical knowledge is not inborn, is a developmental trajectory to be found in contrastive C/V processing?

3. INFANT DEVELOPMENT

3.1. Early Speech Perception

Both C and V contrasts can be discriminated within the first months of life. On the one hand, categorical perception appears early for C (Eimas et al. 1971), but not for V (Swoboda et al. 1976,

Table 1 Crosslinguistic consistency in adult auditory lexical processing[a]

Language	Task(s)	Lexical association	Reference(s)
British English	Word reconstruction	C	van Ooijen (1996)
	Word-spotting	C	Norris et al. (1997)
	Word-spotting	C	Norris et al. (2001)
	Syllable detection	C/V ND	Kearns et al. (2002)
	Primed lexical decision (auditory)	C	Delle Luche et al. (2014)
	Word reconstruction	C	Sharp et al. (2005)
American English	Word learning	C	Creel et al. (2006)
	Segmentation of artificial languages	C/V ND	Newport & Aslin (2004)
	Word-spotting	C	Newman et al. (2011)
	Word reconstruction	C	Moates et al. (2002)
Australian English	Word learning	C	Escudero et al. (2016)
Dutch	Word-spotting	C	McQueen & Cutler (1998)
	Primed lexical decision (auditory)	C/V ND	Cutler et al. (1999)
	Word reconstruction	C	Cutler et al. (2000)
German	Word-spotting	C	Hanuliková et al. (2011)
Spanish	Word reconstruction	C	Cutler et al. (2000)
	Word reconstruction	C	Marks et al. (2002)
French	Word-spotting	C	Dumay et al. (2002)
	Segmentation of artificial languages	C	Bonatti et al. (2005)
	Segmentation of artificial languages	C	Mehler et al. (2006)
	Primed lexical decision (auditory)	C	Delle Luche et al. (2014)
	Word learning	C	Havy et al. (2014)
	Word learning (Cantonese)	C/V ND	Poltrock et al. (forthcoming)
Italian	Segmentation of artificial languages	C	Toro et al. (2008)
Slovak	Word-spotting	C	Hanuliková et al. (2010)
Russian	Word-spotting (controls absent)	C/V ND	Alexeeva et al. (2017)
	Segmentation of artificial languages	C	Gomez et al. (2018)
Sesotho	Word-spotting	C	Cutler et al. (2002a)
Berber	Word-spotting	C/V ND	El Aissati et al. (2012)
Japanese	Word-spotting	C	McQueen et al. (2001)
	Word reconstruction	C	Cutler & Otake (2002)
	Word-spotting	C	Cutler et al. (2009)
Mandarin	Word reconstruction	Initial/final ND	Wiener & Turnbull (2016)
	Segmentation of artificial languages	V + T	Gomez et al. (2018)
	Word learning (Cantonese)	C/V ND	Poltrock et al. (forthcoming)
Cantonese	Segmentation of artificial languages	V + T	Gomez et al. (2018)
	Word learning	C/V ND	Poltrock et al. (forthcoming)
	Morpheme-spotting	C	Yip (2004a,b)

[a]Effects marked "C" show stronger lexical association for C than V. Expected C patterns: In word reconstruction, words are easier to find by V replacement (so, all C stay the same). In word-spotting, a C-only context is treated as part of the same frame, and thus inhibits response, but a context including V can be treated as a separate frame. In artificial language learning and word learning, learning is better with C than with V consistency. In lexical decision: Word recognition is facilitated more by prior presentation of C than of V frame.
Abbreviations: initial/final, initial/final parts of syllables in Chinese phonology; ND, no difference; V + T, tone-carrying vowel.

Cheour-Luhtanen et al. 1995). On the other hand, V may even be perceived and discriminated in utero (Lecanuet et al. 1987, Shahidullah & Hepper 1994, Groome et al. 1997), whereas no such evidence exists for C (for a review, see Granier-Deferre et al. 2011).

Newborns are more sensitive to a V than to a C change. When first habituated with a small set of syllables sharing either V or C, newborns reacted more to an added syllable with a new V than to one with a new C (Bertoncini et al. 1988). In a more recent study using brain imaging with near-infrared spectroscopy (NIRS), newborns were found to better memorize and retain the V than the C of a just-familiarized new word form (Benavides-Varela et al. 2012). Lastly, there seem to be developmental differences in the timing of C and V acquisition, with native V categories being learned earlier (by 6 months; Kuhl et al. 1992, Polka & Werker 1994) than C categories (by 10–12 months; Werker & Tees 1984). A recent study even suggests that crosslinguistic processing differences in the perception of a V contrast may be observed in 1-to-3-day-olds, suggesting rapid pre- or postnatal acquisition of V features (Moon et al. 2013), even though V information is actually less reliable in speech than C information, as described in Section 2.1. For infant-directed speech, Bouchon et al. (2015) used auditorily adjusted frequency measures (mel frequency cepstral coefficients, or MFCCs) to evaluate the degree of acoustic difference in pronunciations of real French-learning infant names and single-feature mispronunciations of those names, for 25 V and 28 C contrasts. The V contrasts proved spectrally less distinct than the C contrasts. Similar results were found in a follow-up study on English (Delle Luche et al. 2017).

Newborns and young infants are highly sensitive to prosodic difference, which concerns the C/V issue in that prosodic variation is principally V-borne. Thus newborns discriminate lists of words varying in stress pattern (Sansavini et al. 1997) or pitch contour (Nazzi et al. 1998b). They also use pitch information differently to group sound sequences depending on whether their mother spoke only French during pregnancy or spoke French together with another language using pitch information more systematically (Abboub et al. 2016). At the sentence level, newborns can discriminate between utterances from different languages as long as the languages differ in their rhythm; where languages belong to the same rhythmic class, there is no discrimination (Nazzi et al. 1998a). Note that rhythm class distinctions can be captured by within-utterance C/V balance (Low et al. 2000; Ramus et al. 1999) or by a language's distribution of C tokens belonging to different C subclasses (Guevara Erra & Gervain 2016).

3.2. The Origin of Infant Processing Biases

The C/V asymmetries have provoked several developmental proposals. Nespor et al. (2003, p. 224) suggested a "division of labor" by which two complementary functional biases could help infants learn language: "[C]onsonants, rather than vowels, are most relevant to build the lexicon, and vowels, rather than consonants, are most relevant for grammatical [and prosodic] information." In the most radical view, these are initial biases: Infants start processing C and V as distinctive linguistic categories from birth (Bonatti et al. 2007, Pons & Toro 2010).

A contrasting proposal invokes learned biases; the distinction emerges during development, as a result of language experience. There are two versions of this proposal, which differ in what exposure is crucial: the lexicon (Keidel et al. 2007) or the acoustic–phonetic properties of the native language (Floccia et al. 2014).

The lexical hypothesis (Keidel et al. 2007) attributes C-advantages in adult listeners' word processing to implicit and acquired knowledge of native language lexical structure. This knowledge fosters preferential processing of C information in lexically related tasks, a preference that varies as the relative informativeness of C and V varies across languages. From a developmental perspective, such distributional knowledge can only be available once a sizeable lexicon has been mastered.

This hypothesis therefore predicts a language-dependent emergence of phonological biases as the lexicon grows during the second year of life.

The acoustic–phonetic hypothesis (Floccia et al. 2014) links the emergence of the infant C-advantage to the acquisition of native-language phonetic categories during the first year of life (Werker & Tees 1984, Kuhl et al. 1992, Polka & Werker 1994). Initially, infants would rely on salient acoustic properties and thus preferentially process sounds that are longer, more periodic, and steadier (Bouchon et al. 2015), that is, vowels. Later, infants would learn that C categories are acoustically more distinct and perceived more categorically than V categories, leading C to become faster and more reliable cues to word identity. Such a switch could also be supported by increases at this time in the fine temporal resolution of the auditory system (Werner et al. 1992). As a result, infants would develop distinct C and V phonological categories, each serving as preferential cues for the processing of different kinds of information (C for the lexicon; V for prosody/syntax). Because this hypothesis relies on phonological acquisition independently of lexical acquisition, it predicts emergence of a C-advantage when phonological acquisition occurs, that is, in the first year of life (earlier than predicted by the lexical hypothesis).

3.3. A Case Study: C-Bias in the Acquisition of French

The first developmental study of C-advantage in lexical processing involved French-learning 20-month-olds (Nazzi 2005). Infants saw triads of novel objects, two of them being assigned the same name and the third a name differing in one phonetic feature. The task required infants to group together the objects with the same name. French-learners successfully learned object labels that differed by one plosive C (e.g., /pize/ versus /tize/) but not by one V (/pize/ versus /pyze/). This pattern appeared as early as 16 months (Havy & Nazzi 2009) and proved independent of C class (holding for nasals, liquids, and fricatives as well as plosives; Nazzi & New 2007) as well as of phoneme position [onset versus coda (Nazzi & Bertoncini 2009) and word-initial vowels (Nazzi & Polka 2018)]. French-learners do eventually use V information to acquire new words, of course, but older children and adults still give greater weight to C than to V information (Havy et al. 2014, Nazzi et al. 2009). Thus, word learning in French shows a consistent C-advantage from 16 months until adulthood.

For familiar word recognition, the C-advantage in French may already be in place by the first birthday. Zesiger & Jöhr (2011) presented French-learning 14-month-olds with images of two objects and examined whether mispronunciations affected looking patterns; C mispronunciation had some effect, but V mispronunciation had none. Poltrock & Nazzi (2015), building on the known effect that 11-month-old infants prefer hearing familiar words over pseudowords (Hallé & de Boysson-Bardies 1996), compared preference at this age for one-feature V versus one-feature C mispronunciations of familiar words, and found that the V mispronunciations were preferred. French-learning infants thus demonstrate a C-advantage in word(-form) recognition even at this early stage of lexical development.

Studies with younger French-learning infants show, however, that this C-advantage is not present in the first months. Bouchon et al. (2015) played 5-month-olds correct and mispronounced versions of their name, and here V but not C mispronunciations impaired name recognition. This evidence of a V-bias is consistent with newborns' higher sensitivity to V than C changes (Benavides-Varela et al. 2012, Bertoncini et al. 1988). Thus French-learning infants' awareness of greater C than V informativeness for lexical processing must come from learning, before 11 but not before 5 months. Nishibayashi & Nazzi (2016) directly investigated this timing issue, using a segmentation task to explore French-learning 6- and 8-month-olds' extraction of word forms from fluent speech. The infants first heard two short stories containing two repeated target words, after which their preference for C versus V mispronunciations of these target words was assessed. At 8 months, infants preferred to listen to V mispronunciations (suggesting that recognition was less affected by the vowel mispronunciation, hence a C-advantage), while the opposite was true for 6-month-olds (hence a V-advantage). This was found for both onset and coda C. The C-advantage at 8 months was confirmed in an ERP study exploring processing of correctly versus mispronounced newly segmented word forms, in which V mispronunciations were processed as target words, while onset C mispronunciations were processed as control words, and coda C mispronunciations showed an intermediate pattern (Von Holzen et al. 2018). The C-advantage for lexical processing in French therefore appears by at least 8 months and is then consistently found until adulthood.

3.4. Cross-Language Variation

The developmental data from French-learning infants do not support the claim of the initial bias hypothesis, according to which the C-advantage should be present from birth. The other claim of the initial bias hypothesis is that the C-advantage should be present crosslinguistically and should not be modulated as a function of native language (especially in early development). To date, relevant data have been collected for Italian, English, Danish, and Mandarin.

Italian data suggest a developmental trajectory similar to that in French. Italian-born newborns show higher sensitivity to V than C changes in a newly familiarized word form, again suggesting no C-advantage at birth (Benavides-Varela et al. 2012). In two related studies, Italian-learning infants were first taught new word form–object pairings (e.g., /kuku/–object A, /dede/–object B) and then tested with word forms made by recombining the C and V (e.g., /keke/) to determine which of the two were more important for word recognition (Hochmann et al. 2011, 2017). Six-month-olds relied mainly on vowels, but 12-month-olds relied mainly on consonants, extending to Italian the pattern of emergence of the C-advantage found in French.

However, studies in Germanic languages suggest that C-bias development may be language dependent. Dutch-learning 9-month-olds are sensitive to C mispronunciations of known words in onset C, but to a lesser extent in C coda (Swingley 2005), as are American English–learners (Jusczyk et al. 1999a). In English, though, mixed results appear across tasks and/or dialectal varieties. When C and V changes were directly compared in the same word-learning study, British English–learning infants at both 16 and 23 months could learn the C- and V-contrasted word pairs equally well; a C-advantage did not appear until 30 months (Floccia et al. 2014, Nazzi et al. 2009). Word recognition studies in the same variety showed greater sensitivity to single-feature C than V mispronunciations of familiar words at 15 months, but similar levels of sensitivity at 12, 18, and 24 months (Mani & Plunkett 2007, 2010). Below 17 months, American English–learning infants do better on V than C minimal pairs in word-learning tasks (Curtin et al. 2009, Stager & Werker 1997, Pater et al. 2004); V minimal pairs are also discriminated by Australian English–learning 15-month-olds (Escudero et al. 2014). Word segmentation, however, showed an adult-like effect in American English–learning 12-month-olds; familiarization to words like *low* or *win* was followed by word-form detection (displayed by longer headturns) in *below* or *window* but not in *slow* or *wind* (Johnson et al. 2003). Just like adults (compare with Section 2.4), these 12-month-olds treated a lone C (*s*, *d*) as part of the ongoing lexical frame, whereas a CV (*be*, *do*) could be treated as a separate frame, freeing the familiarized form for detection. In English, C/V perceptual development may pattern differently across varieties.

Danish offers a particularly interesting case in this debate. First, it is a language that, unusually, has more vowels than consonants (31 vowels, including støds, versus 16 consonants). Second, consonants are often underarticulated or reduced, increasing the relative salience of vowels. These two phenomena should disfavor acquiring a C-advantage if its learning is based on acoustic–phonetic properties. In the first test of this hypothesis (Højen & Nazzi 2016), Danish-learning 20-month-olds who were taught new word pairs differing either by a consonant or by a vowel proved to rely more on vowels than consonants, indicating a V-advantage (alas, comparable adult data for Danish are not yet available).

The newest data come from children learning a tone language. Tone languages are also interesting, because tones are carried mostly by vowels, which might affect the relative weight carried by V and C in lexical processing. These effects could go in either direction. On the one hand, since tones are expressed on V, acoustic V realization will be more variable, which could make V category identification more difficult and lead to an increased C-advantage. On the other hand, the fact that vowels also carry tone information could increase children's attention to them, decreasing or even reversing the C-advantage. Furthermore, many tone languages have heavily constrained syllable phonology (e.g., they prohibit complex onsets and most codas), leading to a lower C-to-V ratio in word structure than the cross-language average, which again could discourage a C-advantage.

Monolingual Mandarin-learning and bilingual Mandarin- and English-learning 24-month-olds reacted in a similar way to C, V, or tone mispronunciations of familiar words in overall looking time on a picture–word task (Wewalaarachchi et al. 2017). But measures of response speed showed group differences. Monolingual toddlers were most sensitive to tone and then to V mispronunciations, and least sensitive to C mispronunciations, while bilingual toddlers were most sensitive to V, then C, and then tone mispronunciations. Toddlers learning Mandarin, either alone or in parallel with English, thus have either no bias or a V-advantage in recognizing familiar words.

Table 2 summarizes 27 infant word-level studies. Clearly the predictions from the initial C-bias proposal are not supported. Newborns do not show a C-bias, and indeed no study has shown a C-bias in infants younger than 8 months. Even clear findings of a first-year C-advantage in lexical processing, from French and Italian, are not replicated in similar studies in other languages. Moreover, neither French nor Italian supports the claim of a C-bias present from birth, as in both languages earlier data showed a V-advantage. Instead, the pattern is consistent with an initial V-advantage but a C-advantage that is learned through development.

3.5. The Balance of Evidence

The infant evidence having failed to support a C-advantage present from birth, what remains to be determined is whether either learned-bias hypothesis better accounts for the attested developmental progression. Insofar as the lexical hypothesis (Keidel et al. 2007) would base the C-advantage on knowledge about native lexical structure, and thus predict emergence only once a sizeable lexicon is acquired, the data do not support it. Two of the French studies directly challenge the lexicon size claim: Nishibayashi & Nazzi's (2016) finding of a C-advantage at 8 months, an age at which infants have, at best, small lexicons, and Poltrock & Nazzi's (2015) observation of the absence of a relationship of lexicon size to C preference in 11-month-olds.

The acoustic–phonetic hypothesis (Floccia et al. 2014), by which emergence of C-advantage is linked to the acquisition of native language acoustic–phonetic properties during the first year of life, is directly supported by the French and Italian evidence of an early developmental switch from a V-advantage in the first 6 months to a C-advantage by 8 months (Benavides-Varela et al. 2012; Bouchon et al. 2015; Hochmann et al. 2011, 2017; Nishibayashi & Nazzi 2016). Yet a pure acoustic–phonetic hypothesis may be too strict, insofar as it is predicated on learning of this speech structure preceding lexical acquisition. Alternative proposals suggest the reverse, namely that prior lexical acquisition enables language-specific phonetic and phonological structure to be acquired (Swingley 2009, Vihman 2017). There is also contrasting evidence that phonological abstraction

Table 2 Direction of C/V-advantage in infant/child lexical processing across languages

Language	Task(s)	Age	Advantage	Reference(s)
French	Discrimination	Newborns	V	Bertoncini et al. (1988)
	Own name recognition	5 months	V	Bouchon et al. (2015)
	Segmentation	6 months	V	Nishibayashi & Nazzi (2016)
		8 months	C	Nishibayashi & Nazzi (2016), Von Holzen et al. (2018)
	Word-form recognition	11 months	C	Poltrock & Nazzi (2015)
	Word recognition	14 months	C	Zesiger & Jöhr (2011)
	Word learning	16 months	C	Havy & Nazzi (2009)
	Word learning/categorization	20 months	C	Nazzi (2005), Nazzi & New (2007), Nazzi & Bertoncini (2009), Nazzi & Polka (2018)
	Word learning/categorization	30 months	C	Nazzi et al. (2009)
	Word learning	3/4/5 years	C	Havy et al. (2014)
Italian	Word-form memory	Newborn	V	Benavides-Varela et al. (2012)
	Word form–object pairing	7 months	V	Hochmann et al. (2017)
		12 months	C	Hochmann et al. (2011)
American English	Word segmentation	12 months	C	Johnson et al. (2003)
	Word learning	14–15 months	V	Stager & Werker (1997), Pater et al. (2004), Curtin et al. (2009)
Australia English	Word learning	15 months	V	Escudero et al. (2014)
British English	Word recognition	12/18/24 months	ND	Mani & Plunkett (2007, 2010)
	Word recognition	15 months	C	Mani & Plunkett (2007)
	Word learning	16/23 months	ND	Floccia et al. (2014)
	Word learning	30 months	C	Nazzi et al. (2009)
Danish	Word learning	20 months	V	Højen & Nazzi (2016)
Mandarin	Word recognition	24 months	V	Wewalaarachchi et al. (2017)
Mandarin and English	Word recognition	24 months	V	Wewalaarachchi et al. (2017)

Abbreviation: ND, no difference.

is possible well before early words are learned (Choi et al. 2017). Most relevantly, there is a strong case for viewing lexical and phonological acquisition in the first year as a continuing and necessary interplay between the two goals (Johnson 2016, Feldman et al. 2013, Werker & Yeung 2005, Yeung & Nazzi 2014).

In fact, each of the proposed hypotheses would seem to have at least some partial credibility. Infants are not equipped with a C-bias at birth, but they are also not unbiased, since a V-advantage at that early stage is supported. The acoustic–phonetic hypothesis has direct support, while the lexical hypothesis cannot find support for the claim that a substantial lexicon is required; but perhaps phonetic structure is best acquired in conjunction with early words, and perhaps even the very first words, learned in the first year, would support a C-advantage. Given that by 6 to 8 months infants start segmenting and memorizing word forms (Jusczyk & Aslin 1995, Saffran et al. 1996, Jusczyk et al. 1999b, Höhle & Weissenborn 2003, Kooijman et al. 2013, Nishibayashi et al. 2015) and even recognize a few words (Tincoff & Jusczyk 1999, 2012; Bergelson & Swingley

2012), analyses of the typicality of this earliest vocabulary must figure on the research agenda (see analyses of normed early lexicons in Hochmann et al. 2011, which suggest that a C-advantage could be learned from these lexicons in both French and Italian). So too should comparisons of C-bias magnitude against individual vocabulary size. Furthermore, as indicated above, cross-language comparisons of detailed phonological structure in early vocabularies and of the detailed phonology of input to infants in their first year are required in order to provide an integrative account of the learning patterns across languages and varieties.

There is also a need for further examination and explanation of what has appeared to be a very early V-advantage. Consider the greater role for V in the perception of longer utterances described in Section 2.1, and consider also that speech from the mother makes a fetus's heart beat faster than other speech does (Kisilevsky et al. 2003), that newborns demonstrate this maternal voice recognition at two days of age (Mehler et al. 1978, DeCasper & Fifer 1980), and that newborns are sensitive to rhythmic structure and possibly to V categories (as described above). Rhythm is dependent on V structure. Stories heard in the womb are preferred by the newborn over unheard stories (DeCasper & Spence 1986). The prebirth listening experience, in the mother's third trimester of pregnancy, can without question support the presence of a V-advantage at an infant's birth.

What does this V-advantage then mean for the infant's early acquisition abilities? One consequence may be that vowel-harmony associations can be learned, even by 7-month-olds acquiring a language without vowel harmony (English; Mintz et al. 2018). Particularly informative here is Hochmann et al.'s (2011, 2017) finding that rule learning by both 6- and 12-month-old Italian-learners drew more on V than on C information. In these studies, infants heard words containing either a C or a V repetition and had to associate the type of repetition with the side on which an object appeared on a screen. Infants could learn the V association at both ages. At 6 months they failed to learn the C association at all, and at 11 months they still gave more weight to V than to C. Developmental changes were also revealed by the fact that while the older infants could learn in this task when presented with both repetitions, the younger infants could learn only when presented with one rule, and failed if given both. Rule learning underlies acquisition at many levels, certainly the acquisition of both phonology and syntax, and is the primary function of V-biases in Nespor et al.'s (2003) proposal. We look forward to such studies being extended to further languages. Prosodic structure across languages relies more on V than C; thus, it would be reasonable to expect a V-bias from birth in all languages (although perhaps modulated by language-specific phonological and prosodic properties).

Is a V-bias primarily of use for accomplishing these learning tasks of the first year, then? Not at all. In adulthood, as discussed in Section 2.1, V carries more speaker-identity information than does C. In infancy this is already useful. Some closely related dialectal varieties differ mainly in their V repertoire, for instance; Catalan and Spanish form one such pair, and infants are able to use this property to distinguish speakers even at 5 months (Ortega-Llebaria & Bosch 2015). At 12 months, both Dutch-learning and Australian English–learning infants extract speaker cues from V independently of linguistic cues (Mulak et al. 2017); interestingly, these infants' performance did not vary directly with the magnitude of the linguistic cues (e.g., F_0, formant structure), suggesting that at this age they were already able to use the speaker-specific cues (voice quality) carried by V. As argued by Polka & Nazzi (forthcoming), the change from V- to C-advantage reflects a shift from a focus on the "who" side of the linguistic message (recognition of known language, familiar voices, infant-directed speech) to a focus on its "what" side (meaning, linked to the onset of lexical acquisition). In positing complementarity of the V- and C-biases, Nespor et al. (2003) made a valuable point. But the learned-bias positions do too. Each bias can

Review in Advance first posted on August 1, 2018. (Changes may still occur before final publication.)

be learned—the V-advantage first, beginning even before birth, and the C-advantage as speech becomes a matter of communication, and of production accompanying perception.

4. CONCLUSION

> . . . apud Hebraeos vocales *literarum animae* appellantur, & literae sine vocalibus *corpora sine anima*. [. . . among the Hebrews, vowels are called the souls of (consonant) letters, and (consonant) letters without vowels are bodies without souls.]
>
> Baruch Spinoza, *Compendium of Hebrew Grammar*, 1677

This survey of the adult and infant processing of vowels and consonants has brought us to the point of view that inspired Spinoza to report this description of the V/C role nearly three-and-a-half centuries ago. Spinoza went on to muse further about the distinction, and offer his own (speech production–based) metaphor of a flute performance, with vowels being the sound created by blowing into the pipe, modulated by alteration of the conduit by fingers blocking the holes (the consonants). Hebrew and other Semitic languages indeed capture the V/C division of labor in their orthography. But even when the different roles are not so visible in written text, many languages use terms for V and C that equally well capture the difference (e.g., soul-letter versus body-letter in Tamil; mother-sound and son-sound in Korean).

We too started with V and C as distinct articulated sound types, and described the acoustic and phonological consequences of this distinction, leading to the implications for vocabularies and hence for adult spoken-language recognition. We then traced the developmental path which enables the adult state. There is much more we could have written, in particular about prosody, and the subtle way in which the two categories influence one another as a function of prosodic structure. We also wish we had been able to call on empirical evidence from more languages. But we hope to have shown, in this first joint summary of the infant and adult evidence on this topic, that the title is accurate: Spoken-language recognition is shaped by the patterning of V and C in speech.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

T.N. acknowledges support from l'Agence Nationale de la Recherche (ANR) Blanche grant (ANR-13-BSH2–0004) and Labex-EFL (ANR-10-LABX-0083), and A.C. acknowledges support from the Australian Research Council (CE140100041). The authors also extend heartfelt thanks to Mark Liberman and *Language Log* commenters for enabling Section 4.

LITERATURE CITED

Abboub N, Nazzi T, Gervain J. 2016. Prosodic grouping at birth. *Brain Lang*. 162:46–59

Alexeeva S, Frolova A, Slioussar N. 2017. Data from Russian help to determine in which languages the Possible Word Constraint applies. *J. Psycholinguist. Res.* 46:629–40

Andrews S. 1996. Lexical retrieval and selection processes: effects of transposed-letter confusability. *J. Mem. Lang.* 35:775–800

Benavides-Varela S, Hochmann J, Macagno F, Nespor M, Mehler J. 2012. Newborn's brain activity signals the origin of word memories. *PNAS* 109:17908–13

Bergelson E, Swingley D. 2012. At 6–9 months, human infants know the meanings of many common nouns. *PNAS* 109:3253–58

Bertoncini J, Bijeljac-Babic R, Jusczyk PW, Kennedy LJ, Mehler J. 1988. An investigation of young infants' perceptual representations of speech sounds. *J. Exp. Psychol.* 117:21–33

Bishop J, Toda K. 2012. Syllabification, sonority, and segmentation: evidence from word-spotting. In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*, pp. 1–15. Chicago: Chicago Linguist. Soc.

Bonatti LL, Peña M, Nespor M, Mehler J. 2005. Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* 16:451–59

Bonatti LL, Peña M, Nespor M, Mehler J. 2007. On consonants, vowels, chickens, and eggs. *Psychol. Sci.* 18:924–25

Bond ZS, Garnes S. 1980. Misperceptions of fluent speech. In *Perception and Production of Fluent Speech*, ed. RA Cole, pp. 115–32. Hillsdale, NJ: Erlbaum

Bouchon C, Floccia C, Fux T, Adda-Decker M, Nazzi T. 2015. Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. *Dev. Sci.* 18:587–98

Caramazza A, Chialant D, Capasso D, Miceli G. 2000. Separable processing of consonants and vowels. *Nature* 403:428–30

Carreiras M, Duñabeitia JA, Molinaro N. 2009. Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming. *Cereb. Cortex* 19:2659–70

Carreiras M, Price CJ. 2008. Brain activation for consonants and vowels. *Cereb. Cortex* 18:1727–35

Cheour-Luhtanen M, Alho K, Kujala T, Sainio K, Reinikainen K, et al. 1995. Mismatch negativity indicates vowel discrimination in newborns. *Hear. Res.* 82:53–58

Choi J, Broersma M, Cutler A. 2017. Early phonology revealed by international adoptees' birth language retention. *PNAS* 114:7307–12

Cole RA, Yan Y, Mak B, Fanty M, Bailey T. 1996. The contribution of consonants versus vowels to word recognition in fluent speech. In *Proceedings of the IEEE International Congress on Acoustics, Speech, and Signal Processing*, pp. 853–56. Piscataway, NJ: IEEE

Costa A, Cutler A, Sebastián-Gallés N. 1998. Effects of phoneme repertoire on phoneme decision. *Percept. Psychophys.* 60:1022–31

Creel SC, Aslin RN, Tanenhaus MK. 2006. Acquiring an artificial lexicon: segment type and order information in early lexical entries. *J. Mem. Lang.* 54:1–19

Curtin S, Fennell C, Escudero P. 2009. Weighting of vowel cues explains patterns of word–object associative learning. *Dev. Sci.* 12:725–31

Cutler A, Chen H-C. 1997. Lexical tone in Cantonese spoken-word processing. *Percept. Psychophys*. 59:165–79

Cutler A, Demuth K, McQueen JM. 2002a. Universality versus language-specificity in listening to running speech. *Psychol. Sci.* 13:258–62

Cutler A, McQueen JM, Jansonius M, Bayerl S. 2002b. The lexical statistics of competitor activation in spoken-word recognition. In *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, pp. 40–45. Canberra: Australas. Speech Sci. Technol. Assoc.

Cutler A, van Ooijen B, Norris D. 1999. Vowels, consonants, and lexical activation. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 3:2053–56. San Francisco: Univ. Calif.

Cutler A, Otake T. 1994. Mora or phoneme? Further evidence for language-specific listening. *J. Mem. Lang.* 33:824–44

Cutler A, Otake T. 2002. Rhythmic categories in spoken-word recognition. *J. Mem. Lang.* 46:296–322

Cutler A, Otake T, McQueen JM. 2009. Vowel devoicing and the perception of spoken Japanese words. *J. Acoust. Soc. Am.* 125:1693–703

Cutler A, Sebastián-Gallés N, Soler-Vilageliu O, van Ooijen B. 2000. Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Mem. Cogn.* 28:746–55

DeCasper AJ, Fifer WP. 1980. Of human bonding: Newborns prefer their mothers' voice. *Science* 208:1174–76

DeCasper AJ, Spence MJ. 1986. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* 9:133–50

Delle Luche C, Floccia C, Granjon L, Nazzi T. 2017. Infants' first words are not phonetically specified: own name recognition in 5-month-old British-English babies. *Infancy* 22:362–88
Delle Luche C, Poltrock S, Goslin J, New B, Floccia C, Nazzi T. 2014. Differential processing of consonants and vowels in the auditory modality: a cross-linguistic study. *J. Mem. Lang.* 72:1–15
Dumay N, Frauenfelder UH, Content A. 2002. The role of the syllable in lexical segmentation in French: word-spotting data. *Brain Lang*. 81:144–61
Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. 1971. Speech perception in infants. *Science* 171:303–6
El Aissati A, McQueen JM, Cutler A. 2012. Finding words in a language that allows words without vowels. *Cognition* 124:79–84
Escudero P, Best CT, Kitamura C, Mulak KE. 2014. Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Front. Psychol.* 5:1059
Escudero P, Mulak KE, Vlach HA. 2016. Cross-situational learning of minimal word pairs. *Cogn. Sci.* 40:455–65
Feldman NH, Myers EB, White KS, Griffiths TL, Morgan JL. 2013. Word-level information influences phonetic learning in adults and infants. *Cognition* 127:427–38
Floccia C, Nazzi T, Luche CD, Poltrock S, Goslin J. 2014. English-learning one- to two-year-olds do not show a consonant bias in word learning. *J. Child Lang.* 41:1085–114
Fogerty D, Humes LE. 2012. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *J. Acoust. Soc. Am.* 131:1490–501
Gomez DM, Mok P, Ordin M, Mehler J, Nespor M. 2018. Statistical speech segmentation in tone languages: the role of lexical tones. *Lang. Speech* 61:84–96
Granier-Deferre C, Ribeiro A, Jacquet A-Y, Bassereau S. 2011. Near-term fetuses process temporal features of speech. *Dev. Sci.* 14:336–52
Groome LJ, Mooney DM, Holland SB, Bentz LS, Atterbury JL. 1997. The heart rate deceleratory response in low-risk human fetuses: effect of stimulus intensity on response topography. *Dev. Psychobiol.* 30:106–13
Guevara Erra R, Gervain J. 2016. The efficient coding of speech: cross-linguistic differences. *PLOS ONE* 11:e0148861
Hallé P, de Boysson-Bardies B. 1996. The format of representation of recognized words in infants' early receptive lexicon. *Infant Behav. Dev.* 19:463–81
Hanuliková A, McQueen JM, Mitterer H. 2010. Possible words and fixed stress in the segmentation of Slovak speech. *Q. J. Exp. Psychol.* 63:555–79
Hanuliková A, Mitterer H, McQueen JM. 2011. Effects of first and second language on segmentation of non-native speech. *Biling. Lang. Cogn.* 14:506–21
Havy M, Nazzi T. 2009. Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy* 14:439–56
Havy M, Serres J, Nazzi T. 2014. A consonant/vowel asymmetry in word-form processing: evidence in childhood and in adulthood. *Lang. Speech* 57:254–81
Hochmann JR, Benavides-Varela S, Flo A, Nespor M, Mehler J. 2017. Bias for vocalic over consonantal information in 6-month-olds. *Infancy* 23:136–51
Hochmann JR, Benavides-Varela S, Nespor M, Mehler J. 2011. Consonants and vowels: different roles in early language acquisition. *Dev. Sci.* 14:1445–58
Höhle B, Weissenborn J. 2003. German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Dev. Sci.* 6:122–27
Højen A, Nazzi T. 2016. Vowel bias in Danish word-learning: Processing biases are language-specific. *Dev. Sci.* 19:41–49
Johnson EK. 2016. Constructing a proto-lexicon: an integrative view of infant language development. *Annu. Rev. Linguist.* 2:391–412
Johnson EK, Jusczyk PW, Cutler A, Norris D. 2003. Lexical viability constraints on speech segmentation by infants. *Cogn. Psychol.* 46:65–97
Johnson K, Ladefoged P, Lindau M. 1993. Individual differences in vowel production. *J. Acoust. Soc. Am.* 94:701–14
Jusczyk PW, Aslin RN. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29:1–23

Jusczyk PW, Goodman MB, Baumann A. 1999a. Nine-month-olds' attention to sound similarities in syllables. *J. Mem. Lang.* 40:62–82

Jusczyk PW, Houston DM, Newsome M. 1999b. The beginning of word segmentation in English-learning infants. *Cogn. Psychol.* 39:159–207

Kearns RK, Norris D, Cutler A. 2002. Syllable processing in English. In *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 1657–60. Baixas, Fr.: Int. Speech Commun. Assoc.

Keidel JL, Jenison RL, Kluender KR, Seidenberg MS. 2007. Does grammar constrain statistical learning? *Psychol. Sci.* 18:922–23

Kewley-Port DT, Burkle TZ, Lee JH. 2007. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 122:2365–75

Kisilevsky BS, Hains SMJ, Lee K, Xie X, Huang H, et al. 2003. Effects of experience on fetal voice recognition. *Psychol. Sci.* 14:220–24

Kolinsky R, Lidji P, Peretz I, Besson M, Morais J. 2009. Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition* 112:1–20

Kooijman V, Junge C, Johnson EK, Hagoort P, Cutler A. 2013. Predictive brain signals of linguistic development. *Front. Psychol.* 4:25

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606–8

Ladefoged P, Disner SF. 2012. *Vowels and Consonants*. Chichester, UK: Wiley-Blackwell. 3rd ed.

Lecanuet J-P, Granier-Deferre C, DeCasper AJ, Maugeais R, Andrieu AJ, Busnel M-C. 1987. Perception et discrimination foetales de stimuli langagiers; mise en évidence à partir de la réactivité cardiaque, résultats préliminaires. *C. R. Acad. Sci.* 305:161–64

Lee H-W, Rayner K, Pollatsek A. 2002. The processing of consonants and vowels in reading: evidence from the fast priming paradigm. *Psychon. Bull. Rev.* 9:766–72

Low EL, Grabe E, Nolan F. 2000. Quantitative characterization of speech rhythm: syllable-timing in Singapore English. *Lang. Speech* 43:377–401

Lupker SJ, Perea M, Davis CJ. 2008. Transposed-letter effects: consonants, vowels and letter frequency. *Lang. Cogn. Process.* 23:93–116

Maddieson I. 1984. *Patterns of Sounds*. Cambridge, UK: Cambridge Univ. Press

Mani N, Plunkett K. 2007. Phonological specificity of vowels and consonants in early lexical representations. *J. Mem. Lang.* 57:252–72

Mani N, Plunkett K. 2010. Twelve-month-olds know their cups from their keps and tups. *Infancy* 15:445–70

Marks EA, Moates DR, Bond ZS, Vazquez L. 2002. Vowel mutability: the case of monolingual Spanish listeners and bilingual Spanish–English listeners. *Southwest J. Linguist.* 21:73–99

McCarthy JJ, Prince AS. 1995. Prosodic morphology. In *Handbook of Phonology*, ed. J Goldsmith, pp. 318–66. Oxford, UK: Blackwell

McQueen JM, Cutler A. 1998. Spotting (different kinds of) words in (different kinds of) context. In *Proceedings of the 5th International Conference on Spoken Language Processing*, 6:2791–94. Canberra: Australas. Speech Sci. Technol. Assoc.

McQueen JM, Otake T, Cutler A. 2001. Rhythmic cues and possible-word constraints in Japanese speech segmentation. *J. Mem. Lang.* 45:103–32

Mehler J, Bertoncini J, Barriere M, Jassik-Gerschenfeld D. 1978. Infant recognition of mother's voice. *Perception* 7:491–97

Mehler J, Peña M, Nespor M, Bonatti LL. 2006. The "soul" of language does not use statistics: reflections on vowels and consonants. *Cortex* 42:846–54

Mintz TH, Walker RL, Welday A, Kidd C. 2018. Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition* 171:95–107

Moates DR, Bond ZS, Stockmal V. 2002. Phoneme frequency in spoken word reconstruction. In *Laboratory Phonology VII*, ed. C Gussenhoven, N Warner, pp. 141–69. The Hague: Mouton

Moates DR, Marks EA. 2012. Vowel mutability in print in English and Spanish. *Ment. Lexicon* 7:326–50

Monaghan P, Shillcock RC. 2003. Connectionist modelling of the separable processing of consonants and vowels. *Brain Lang.* 86:83–98

Moon C, Lagercrantz H, Kuhl PK. 2013. Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatr*. 102:156–60

Mueller JL, Bahlmann J, Friederici AD. 2008. The role of pause cues in language learning: the emergence of event-related potentials related to sequence processing. *J. Cogn. Neurosci.* 20:892–905

Mulak KE, Bonn CD, Chládková K, Aslin RN, Escudero P. 2017. Indexical and linguistic processing by 12-month-olds: discrimination of speaker, accent and vowel differences. *PLOS ONE* 12:e0176762

Nazzi T. 2005. Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition* 98:13–30

Nazzi T, Bertoncini J. 2009. Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Lang. Speech* 52:463–80

Nazzi T, Bertoncini J, Mehler J. 1998a. Language discrimination by newborns: towards an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24:756–66

Nazzi T, Floccia C, Bertoncini J. 1998b. Discrimination of pitch contours by neonates. *Infant Behav. Dev.* 21:779–84

Nazzi T, Floccia C, Moquet B, Butler J. 2009. Bias for consonantal information over vocalic information in 30-month-olds: cross-linguistic evidence from French and English. *J. Exp. Child Psychol.* 102:522–37

Nazzi T, New B. 2007. Beyond stop consonants: consonantal specificity in early lexical acquisition. *Cogn. Dev.* 22:271–79

Nazzi T, Polka L. 2018. The consonant bias in word learning is not determined by position within the word: evidence from vowel-initial words. *J. Exp. Child Psychol.* 174:103–11

Nespor M, Peña M, Mehler J. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue Linguaggio* 2:203–30

Nettle D. 1994. A behavioural correlate of phonological structure. *Lang. Speech* 37:425–29

New B, Araújo V, Nazzi T. 2008. Differential processing of consonants and vowels in lexical access through reading. *Psychol. Sci.* 19:1223–27

New B, Nazzi T. 2014. The time course of consonant and vowel processing during word recognition. *Lang. Cogn. Neurosci.* 29:147–57

Newman RS, Sawusch JR, Wunnenberg T. 2011. Cues and cue interactions in segmenting words in fluent speech. *J. Mem. Lang.* 64:460–76

Newport EL, Aslin RN. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48:127–62

Nishibayashi L-L, Goyet L, Nazzi T. 2015. Early speech segmentation in French-learning infants: monosyllabic words versus embedded syllables. *Lang. Speech* 58:334–50

Nishibayashi L-L, Nazzi T. 2016. Vowels, then consonants: early bias switch in recognizing segmented word forms. *Cognition* 155:188–203

Nooteboom SG, Doodeman GJN. 1980. Production and perception of vowel length in spoken sentences. *J. Acoust. Soc. Am.* 67:276–87

Norris D, McQueen JM, Cutler A, Butterfield S. 1997. The Possible-Word Constraint in the segmentation of continuous speech. *Cogn. Psychol.* 34:191–243

Norris D, McQueen JM, Cutler A, Butterfield S, Kearns R. 2001. Language-universal constraints on speech segmentation. *Lang. Cogn. Process.* 16:637–60

Onnis L, Monaghan P, Christiansen MH, Chater N. 2004. Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, ed. K Forbus, D Gentner, T Regier, pp. 1047–52. Austin, TX: Cogn. Sci. Soc.

Onnis L, Monaghan P, Richmond K, Chater N. 2005. Phonology impacts segmentation in speech processing. *J. Mem. Lang.* 53:225–37

Ortega-Llebaria M, Bosch L. 2015. Cues to dialectal discrimination in early infancy: a look at prosodic, rhythmic and segmental properties. In *The Phonetics–Phonology Interface: Representations and Methodologies*, ed. J Romero, M Riera, pp. 55–71. Amsterdam: Benjamins

Owren MJ, Cardillo GC. 2006. The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *J. Acoust. Soc. Am.* 119:1727–39

Parikh G, Loizou P. 2005. The influence of noise on vowel and consonant cues. *J. Acoust. Soc. Am.* 118:3874–88
Pater J, Stager C, Werker JF. 2004. The perceptual acquisition of phonological contrasts. *Language* 80:384–402
Peña M, Bonatti LL, Nespor M, Mehler J. 2002. Signal-driven computations in speech processing. *Science* 298:604–7
Perea M, Acha J. 2009. Does letter position coding depend on consonant/vowel status? Evidence with the masked priming technique. *Acta Psychol*. 130:127–37
Perea M, Carreiras M. 2008. Do orthotactics and phonology constrain the transposed-letter effect? *Lang. Cogn. Process.* 23:69–92
Perruchet P, Peereman R, Tyler MD. 2006. Do we need algebraic-like computations? A reply to Bonatti, Peña, Nespor, and Mehler 2006. *J. Exp. Psychol. Gen.* 135:322–26
Perruchet P, Tyler MD, Galland N, Peereman R. 2004. Learning non-adjacent dependencies: no need for algebraic-like computations. *J. Exp. Psychol. Gen.* 133:573–83
Pisoni DB. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* 13:253–60
Polka L, Nazzi T. Interacting processes and developmental biases allow learners to crack the "what" code and the "who" code in spoken language. *Appl. Psycholinguist.* Forthcoming. **http://doi.org/10.1017/S0142716418000218**
Polka L, Werker JF. 1994. Developmental changes in perception of non-native vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* 20:421–35
Poltrock S, Chen H, Kwok C, Cheung H, Nazzi T. Adult learning of words in a nonnative language: consonants, vowels, and tones. *Front. Psychol.* Forthcoming. **http://doi.org/10.3389/fpsyg.2018.01211**
Poltrock S, Nazzi T. 2015. Consonant/vowel asymmetry in early word form recognition. *J. Exp. Child Psychol.* 131:135–48
Pons F, Toro JM. 2010. Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition* 116:361–67
Ramus F, Nespor M, Mehler J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73:265–92
Repp BH. 1984. Categorical perception: issues, methods, findings. In *Speech and Language*, vol. 10: *Advances in Basic Research and Practice*, ed. NJ Lass, pp. 243–335. New York: Academic
Ridouane R. 2008. Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber. *Phonology* 25:321–59
Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–28
Sansavini A, Bertoncini J, Giovanelli G. 1997. Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33:3–11
Schubert T, Kinoshita S, Norris D. 2018. What causes the greater perceived similarity of consonant-transposed nonwords? *Q. J. Exp. Psychol.* 71:642–56
Seidenberg MS, MacDonald MC, Saffran JR. 2002. Does grammar start where statistics stop? *Science* 298:553–54
Shahidullah S, Hepper PG. 1994. Frequency discrimination by the fetus. *Early Hum. Dev.* 36:13–26
Shankweiler D, Studdert-Kennedy M. 1967. Identification of consonants and vowels presented to left and right ears. *Q. J. Exp. Psychol.* 19:59–63
Sharp DJ, Scott SK, Cutler A, Wise RJS. 2005. Lexical retrieval constrained by sound structure: the role of the left inferior frontal gyrus. *Brain Lang*. 92:309–19
Soto-Faraco S, Sebastián-Gallés N, Cutler A. 2001. Segmental and suprasegmental mismatch in lexical access. *J. Mem. Lang.* 45:412–32
Stager CL, Werker JF. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388:381–82
Swingley D. 2005. 11-month-olds' knowledge of how familiar words sound. *Dev. Sci.* 8:432–43
Swingley D. 2009. Contributions of infant word learning to language development. *Philos. Trans. B* 364:3617–32
Swinney DA, Prather P. 1980. Phoneme identification in a phoneme-monitoring experiment: the variable role of uncertainty about vowel contexts. *Percept. Psychophys.* 27:104–10
Swoboda P, Morse PA, Leavitt LA. 1976. Continuous vowel discrimination in normal and at-risk infants. *Child Dev*. 47:459–65

Taft M, Chen H-C. 1992. Judging homophony in Chinese: the influence of tones. In *Language Processing in Chinese*, ed. H-C Chen, OJL Tzeng, pp. 151–72. Amsterdam: North-Holland

Tincoff R, Jusczyk PW. 1999. Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* 10:172–75

Tincoff R, Jusczyk PW. 2012. Six-month-olds comprehend words that refer to parts of the body. *Infancy* 17:432–44

Toro JM, Nespor M, Mehler J, Bonatti L. 2008. Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychol. Sci.* 19:137–44

Tsang KK, Hoosain R. 1979. Segmental phonemes and tonal phonemes in comprehension of Cantonese. *Psychologia* 22:222–24

van Ooijen B. 1996. Vowel mutability and lexical selection in English: evidence from a word reconstruction task. *Mem. Cogn.* 24:573–83

van Ooijen B, Cutler A, Norris D. 1991. Detection times for vowels versus consonants. In *Proceedings of EUROSPEECH 91*, 3:1451–54. Baixas, Fr.: Int. Speech Commun. Assoc.

Vihman MM. 2017. Learning words and learning sounds: advances in language development. *Br. J. Psychol.* 108:1–27

Von Holzen K, Nishibayashi L-L, Nazzi T. 2018. Word segmentation abilities: an ERP study. *Brain Sci*. 8:24

Werker JF, Tees RC. 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7:49–63

Werker JF, Yeung HH. 2005. Infant speech perception bootstraps word learning. *Trends Cogn. Sci.* 9:519–27

Werner LA, Marean GC, Halpin CF, Spetner NB, Gillenwater JM. 1992. Infant auditory temporal acuity: gap detection. *Child Dev*. 63:260–72

Wewalaarachchi TD, Wong LH, Singh L. 2017. Vowels, consonants, and lexical tones: sensitivity to phonological variation in monolingual Mandarin and bilingual English–Mandarin toddlers. *J. Exp. Child Psychol.* 159:16–33

Wiener S, Turnbull R. 2016. Constraints of tones, vowels, and consonants on lexical selection in Mandarin Chinese. *Lang. Speech* 59:59–82

Wood CC, Day RS. 1975. Failure of selective attention to phonetic segments in consonant–vowel syllables. *Percept. Psychophys.* 17:346–50

Ye Y, Connine CM. 1999. Processing spoken Chinese: the role of tone information. *Lang. Cogn. Process.* 14:609–30

Yeung HH, Nazzi T. 2014. Infants can generalize phonetic contrasts learned from ostensive and referential labels. *Cognition* 132:151–63

Yip MCW. 2004a. Possible-word constraints in Cantonese speech segmentation. *J. Psycholinguist. Res.* 33:165–73

Yip MCW. 2004b. Interference effects of possible-word constraints (PWC) in Cantonese speech segmentation. *Psychologia* 47:169–77

Zesiger P, Jöhr J. 2011. Les représentations phonologiques des mots chez le jeune enfant [Phonological representations of words in the young child]. *Enfance* 3:293–309

Zwitserlood P. 1996. Form priming. *Lang. Cogn. Process.* 10:121–36