Canadian Journal of Linguistics/Revue canadienne de linguistique, 67(4): 670–682, 2022
doi: 10.1017/cnj.2022.36


# *Underspecification in time*

WILLIAM IDSARDI
*University of Maryland, College Park, USA*
idsardi@umd.edu

***Abstract***

Substance-free phonology or SFP (Reiss 2017) has renewed interest in the question of abstraction in phonology. Perhaps the most common form of abstraction through the absence of substance is underspecification, where some aspects of speech lack representation in memorized representations, within the phonology or in the phonetic implementation (Archangeli 1988, Keating 1988, Lahiri and Reetz 2010 among many others). The fundamental basis for phonology is argued to be a mental model of speech events in time, following Raimy (2000) and Papillon (2020). Each event can have properties (one-place predicates that are true of the event), which include the usual phonological features, and also structural entities for extended events like moras and syllables. Features can be bound together in an event, yielding segment-like properties. Pairs of events can be ordered in time by the temporal logic precedence relation represented by '<'. Events, features and precedence form a directed multigraph structure with edges in the graph interpreted as "maybe next". Some infant bimodal speech perception results are examined using this framework, arguing for underspecification in time in the developing phonological representations.

**Keywords:** phonology, speech perception, speech development, time, underspecification

***Résumé***

La phonologie sans substance ou SFP (Reiss 2017) a renouvelé l'intérêt pour la question de l'abstraction en phonologie. La forme la plus commune d'abstraction par l'absence de substance est peut-être la sous-spécification, où certains aspects de la parole manquent de représentation dans les représentations mémorisées, au sein de la phonologie ou dans l'implémentation phonétique (Archangeli 1988, Keating 1988, Lahiri and Reetz 2010 parmi beaucoup d'autres). La base fondamentale de la phonologie est un modèle mental des événements de la parole dans le temps, suivant Raimy (2000) et Papillon (2020). Chaque événement peut avoir des propriétés (prédicats monovalents qui sont vrais pour l'événement) qui incluent les traits phonologiques classiques, et aussi des entités structurelles pour les événements étendus comme les mores et les syllabes. Les traits peuvent être liés entre eux dans un événement, ce qui donne des propriétés de type segment. Les paires d'événements peuvent être ordonnées dans le temps par la relation logique temporelle de précédence, soit <. Les événements, les traits et la précédence forment une structure multigraphique dirigée, les arêtes du graphe étant interprétées comme « peut-être le prochain ». Certains résultats de la perception bimodale de la parole chez le nourrisson sont examinés à l'aide de ce cadre, ce qui

plaide en faveur d'une sous-spécification temporelle dans les représentations phonologiques en développement.

**Mots clés:** phonologie, perception de la parole, développement de la parole, temps, sous-spécification

## 1. Introduction

Substance-free phonology or SFP (Reiss 2017) has renewed interest in the question of abstraction in phonology. Perhaps the most common form of abstraction through the absence of substance is underspecification, where some aspects of speech lack representation in memorized representations, within the phonology or in the phonetic implementation (Archangeli 1988, Keating 1988, Lahiri and Reetz 2010 among many others). The main locus of underspecification has been features within segments, but in this article I will build toward an argument saying that analyses should also allow for the underspecification of temporal order between items – that not all elements have a definite order established between them at all levels of representation. To do this, I will begin with a proposal for a general framework for phonological representations in terms of directed graphs of phonological events (the nodes or vertices of the graph) where features are monadic properties (one-place predicates) of events and where pairs of events can stand in a dyadic relation (two-place predicate) of temporal precedence (the directed edges of the graph). Modalities other than speech which make use of spatial relationships (sign languages, Braille, etc.) will also need further spatial relations in addition to the temporal precedence relation. It is also possible that additional spatial relations would be useful for speech, but the ability to perceive speech from monaural recordings suggests that spatial localization is not a significant factor in speech perception, whereas the location of a handshape can be contrastive in sign language (Brentari 1998). Importantly, this proposal allows for underspecification of features (that is, not all events have all monadic properties) and of time (that is, not all pairs of events have a dyadic relation of precedence).

I am completely in favour of abstract phonology, that is, phonology not completely determined by the phonetics, in which there are interesting "slippages" between the representations for speech in the memory, action and perception systems. This is the position taken in Avery and Idsardi (2001), for instance. However, in order to be usable, any model has to have quasi-veridical mappings between the systems, that is, constrained variation in the mapping between auditory cues and phonological categories, for instance. So, one understanding of SFP would be that these relationships are instead unconstrained and learned in all cases. (Most of the discussion of this issue has centred on phonological features, whereas this article will focus instead on time and temporal relationships.) For features, results from neonates and young infants have consistently revealed various biases in the systems relating memory, action and perception, and I will discuss one of these studies (Baier et al. 2007), below. Idemaru and Holt (2014) also provide extremely relevant evidence for constrained variation, in two ways. First, voice onset time

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

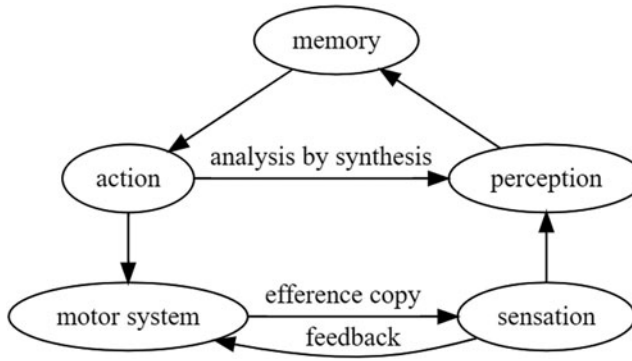


**Figure 1:** A memory-action-perception loop.

(VOT) serves as the primary cue for English stops, with F0 on the following vowel serving only as a secondary cue. Second, while it is possible, under brief exposures to a novel idiolect, to reduce the reliance on the secondary cue, and even extinguish it – and therefore this cue is, to a certain extent, malleable and plastic –, in Idemaru and Holt's experiments, it was not possible to reverse the relationship between F0 and category membership. That is, higher F0 on the following vowel cannot be made to signal "voicing", nor lower F0 signal "voicelessness".

If phonological category formation is as labile and arbitrary as SFP proponents seem to suggest (contra studies like Heffner et al. 2019), then one potentially fruitful area of research would be the influence of orthography on speech production and perception in literate speakers. There certainly have been claims for (Rastle et al. and Davis 2011) and against (Damian and Bowers 2009, Mitterer and Reinisch 2015) orthographic effects in speech production and perception tasks not involving overt use of orthography. (There is robust evidence for orthographic effects in tasks that do overtly involve orthography, e.g., Li et al. 2015.)

## 2. Events, features and precedence

In my view, phonology must interface effectively between at least three components of the mind/brain: *memory* (long-term memory for the lexicon, short-term memory for what Baddeley (1986) has termed the 'phonological loop', and episodic memory for particular people and situations), *action* (for the motor control of the oral, manual and other articulators) and *perception* (at least auditory, visual, and tactile), which together constitute a memory-action-perception loop (Poeppel and Idsardi 2011). A small portion of this system is shown schematically in Figure 1.

In speech production, items are drawn from long-term memory and assembled into a motor plan to be executed by the motor system (downwards along the left side of Figure 1). In speech perception, external sensations are perceived and ultimately recognized as conglomerations of items in long-term memory (upwards along the right side of Figure 1). The motor system is coupled to sensation with efference

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

copies and feedback (see Gallistel 1980 for a review), and a similar, more abstract relation between action and perception has been proposed in the analysis-by-synthesis model of speech perception (Halle and Stevens 1963, Bever and Poeppel 2010).

I will argue here for a view that part of the fundamental basis for phonology is a mental model of individuated speech events in time, following Raimy (2000) and Papillon (2020) (and also strongly influenced by Pietroski 2005, 2018). Each event can have properties (monadic predicates that are true of that event), which we will equate with the usual phonetic features **and** with structural entities like moras and syllables. The idea is that moras and syllables cover intervals of time in a continuous model, so these intervals are then single events in the discrete model, which are in parallel temporally with the segmental events that are "contained" in the intervals for the moras or syllables. Following Jakobson et al. (1951), the features provide the "glue" between action and perception (because features must have definitions in terms of both action/articulation and perception/audition), and I also claim, following Poeppel et al. (2008) (but with much less evidence), that features constitute a basic element of storage in long-term memory. An individual event can have multiple features; equivalently, features can be bound together in an event, yielding segment-like properties (like Firthian sounds; see Firth 1948, Kazanina et al. 2017), or events can contain relatively few features, even just a single feature, leading to autosegmental-like behaviours (akin to Firthian prosodies).

The other necessary concept is at least one temporal relation between events: precedence. For sign languages, the events are also located in signing space, and so we will require spatial relations in addition to the temporal relation. Similarly, orthographic systems will require a specification of the relevant spatial relations, which may require more than one spatial dimension, as is the case for Hangul, and a mapping between spatial and temporal relations, as in Braille reading where the fingertip scans an embossed line of text. The kinds of sensible temporal relations depend greatly on the model of time that is employed in the model. One explicit model is a continuous timeline, as employed in Articulatory Phonology (Browman and Goldstein 1989) and Time Map Phonology (Carson-Berndsen 1998); see also Bird and Klein (1990). Events in such models are construed as continuous intervals of time, and admit for temporal relations such as overlap in time (sharing some sub-interval of the timeline), precedence (i.e., notions of before and after), and whether events have a functional form (as in Articulatory Phonology, or in the musical tradition of analysis into attack, sustain and release; see Françoise et al. 2012), and higher-order temporal relations like phasing. These ideas also pervade continuous signal processing models, especially continuous wavelet models which provide an event-based alternative to Fourier analysis (e.g., Suni et al. 2017). The continuous-time gestural score for "palm" from Browman and Goldstein (1989: 212) is shown in Figure 2.

It is commonly observed that phonetic parameters in articulation and audition can have a wide range of continuous values, whereas phonological representations typically quantize (or discretize) these values into a small number of options, and also reduce the dimensionality of the problem by picking a primary dimension for the phonological parameter, or by grouping together several phonetic dimensions

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

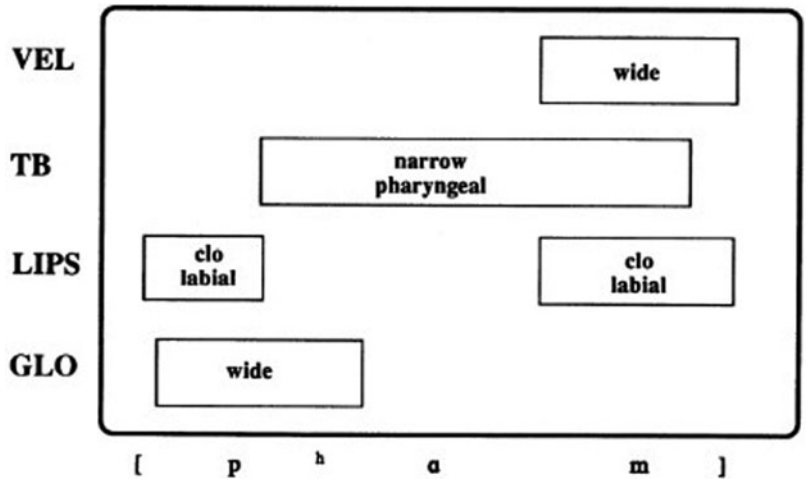


**Figure 2:** Gestural score for "palm", Browman and Goldstein 1989: 212.

through the use of phonological latent variables. These techniques of discretization, dimensionality reduction and latent variables are also common techniques in machine learning more generally (Murphy 2012). Browman and Goldstein (1989) also provide a discrete version of their gestural score, in what they term "point notation". A point notation diagram, again for "palm", is shown in Figure 3. Browman and Goldstein also endorse, at least implicitly, a mapping between Figure 2 and Figure 3, that is, between continuous and discrete representations.

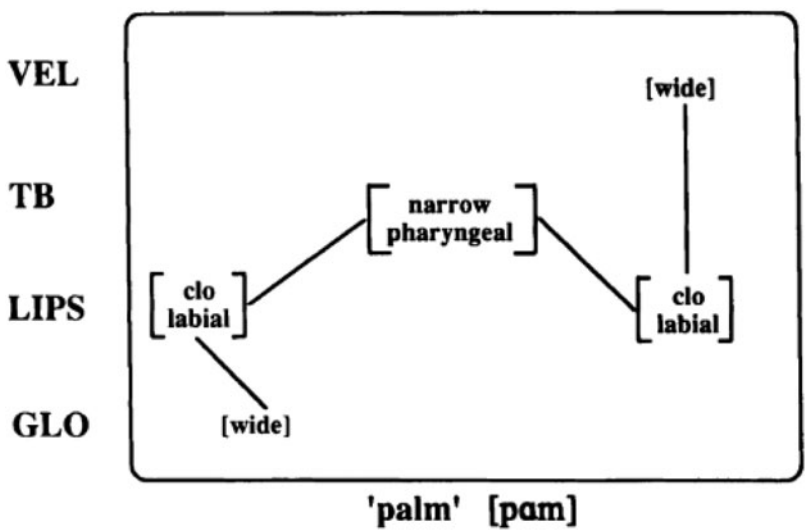


**Figure 3:** Point notation diagram for "palm", Browman and Goldstein 1989: 213.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

Generative phonology has also taken a much more discrete approach to phonological representations, with features and segments conceived as points in time, albeit with the temporal aspects usually more implicit than explicit, indicated by position in a string, or by spatial position in a diagram. The relevant, but often vague, temporal notions have been something like a containment relation (two features contained in the same segment or two segments contained in an onset), and a synchronization relation in the association lines of Autosegmental Phonology (Goldsmith 1976, Sagey 1988, Coleman and Local 1991).

There is not sufficient space in this article to discuss the general issue of continuous and discrete models of time, or physical models of time, which are difficult and contentious issues in physics and metaphysics. Moreover, whatever the physical nature of time, human mental models of time are probably only quasi-veridical, agreeing in some ways with the physical world, but failing to be fully veridical with it (and hence *models*). The idea that perception in part chunks the perceived world into discretized pieces has a long history, and much recent support in multiple time-scale models for speech (Luo and Poeppel 2012, Teng et al. 2016). I find this view compelling, but it does not entail that discrete representations are the *only* data structures used by the mind/brain for storing, acting and perceiving speech. That is, it is certainly possible, even likely, that there are mappings between continuous and discrete models in terms of discretization (continuous to discrete in perception) and production (as in the motor articulation of an event in terms of prolonged periods for the attack, sustain and release) akin to those in Browman and Goldstein (1989).

Importantly, most of these models assume a linear notion of time (implied by the term *timeline*). However, following Raimy (2000) and Papillon (2020), the view adopted here is that phonological time can have loops, and that the appropriate representation for temporal order is a **directed graph** (or, more properly, a multigraph, see, e.g., Chartrand and Zhang 2012), which is allowed to have cycles, which are used to analyze repeated event sequences such as reduplication in language, and repetition in music (Bamberger 1991, Idsardi and Raimy 2008). There is one large difference between Raimy (2000) and Papillon (2020): Raimy (2000) retains the framework of Autosegmental Phonology and association lines, whereas Papillon (2020) abandons association lines in favour of a network flow model of simultaneous articulations, compare Figure 4a and Figure 4b.

These are examples of the overapplication of progressive nasalization in reduplication in forms such as [aŋẽn] => [ãŋẽn-ãŋẽn]. In both cases [+nasal] comes to have "scope" over the entire form. This is handled in Raimy (2000) by association lines between [+nasal] and the individual segment-sized events; in comparison, Papillon (2020) handles such cases by placing the [+nasal] event in parallel to the entire stream of segmental events.

I will adopt here Papillon's conjecture that a single temporal relation suffices, and that this is the precedence relation, which is the basic relation in discrete temporal logic (van Benthem 1983). The association lines in Autosegmental Phonology – interpreted as a synchronization relation between events – is an obvious possible addition to the precedence relation, but Papillon's model employs a strategy of reducing

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

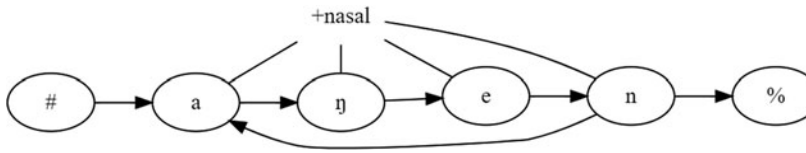


**Figure 4a.** [ãŋẽn-ãŋẽn] Raimy 2000: 18.

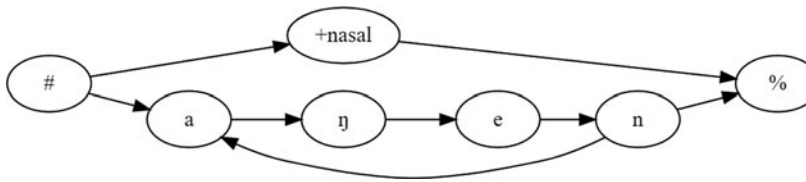


**Figure 4b:** [ãŋẽn-ãŋẽn] reinterpreted via Papillon 2020.

synchronization effects to a lack of precedence specification between events. In the discrete representations, the directed edges can also be interpreted as something akin to "maybe next" (◇**N** in modal temporal logic, Goranko and Rumberg 2020; there is no simple successor relation in continuous time). Directed edges can occur both within a "tier" or "stream" of events, and across tiers/streams. This has the consequence that the production of a motor plan from a phonological graph is not (as in Raimy 2000) finding a path in a graph from beginning (#) to end (%), but rather finding the flow through a network (Barabási 2016), because parallel edges can be traversed simultaneously (e.g., [+nasal] in Figure 4b).

In summary, the present proposal is that spoken language has an <E, F, P> structure for phonological representations, consisting of a set of events (E, abstract points in time), a set of monadic predicates over events (F, features, e.g., a feature f could be true of event e, f(e)), and a (potentially incomplete) binary relation, <, of temporal precedence between events (such that e1 < e2 is true if e1 happens before e2). I will call this the EFP model.

This sketch has been rather abstract and terse, but I hope that the small case study that is examined in the following sections will make the relevant points clearer.

## 3. A PUZZLE IN EARLY INFANT BIMODAL SPEECH PERCEPTION

Infants can show a puzzling range of phonetic abilities and deficits in comparison with adults, out-performing adults on many phonetic perception tasks while lagging behind in other ways. Some null results using one procedure can be overturned with more sensitive procedures and some contrasts are "better" than others in terms of effect size and various acoustic or auditory measures of similarity (Sundara et al. 2018). There are additional oddities in the infant speech perception literature, including the fact that the syllabic stimuli generally need to be much longer than the average syllable durations in adult speech (often twice as long). One persistent idea is that infants start with a syllable-oriented perspective and

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

only later move to a more segment-oriented one (Bertoncini and Mehler 1981, MacNeilage 2008), and that in some languages adults still have a primary orientation for syllables, at least for some speech production tasks (O'Seaghdha et al. 2010, though see various replies, e.g., Qu et al. 2012, Li et al. 2015).

Baier et al. (2007) investigated audio-visual (AV) integration in two-month-old infants, replicating and extending the results of Kuhl and Meltzoff (1982) and Patterson and Werker (2003) on bimodal speech perception in infants. Infants were presented with a single audio track along with two synchronized silent videos of the same person (Rebecca Baier) articulating single syllables, presented on a large TV screen, and the infants' looks to each face were tabulated and analyzed. The infants were able to visually distinguish and match articulating faces with synchronized audio for [a], [i] and [u], replicating the previous findings. Taking this one step further, the infants could also detect dynamic syllables, matching faces with audio when tested on [wi] vs. [i]. However, we were subsequently puzzled by infants' inability to distinguish between [wi] and [ju] which, moreover, correspond to the very high frequency English words "we" and "you". The overall proportion of matching looks was 0.52 for [wi] ($n = 20$, $p > 0.5$, n.s.) and 0.49 for [ju] ($n = 18$, $p > 0.5$, n.s.); neither was significantly different from chance, and they were not different from each other (unpaired $n = 21$, $p > 0.69$ n.s., paired $n = 17$, $p > 0.71$ n.s., including only those infants who completed both conditions). Furthermore, when they were presented with [ju] audio alongside [wi] and [i] faces, they matched the [ju] audio with the [wi] video. These behaviours are at least consistent with a syllable-oriented point of view in terms of features: infants hear a dynamic syllable with something [round] and something [front] in it, but they cannot tell the relative order of the [front] and [round] events. This is also consistent with the relatively poor abilities of infants to detect differences in serial order (Lewkowicz 2004, Warren 2008). Importantly, this is not to say that infants cannot hear a difference between [wi] and [ju], which is strongly signaled acoustically by the trajectory of the second formant: rising F2 in [wi] and falling F2 in [ju]. I am confident that dishabituation experiments on infants would succeed on this contrast, which is a relatively easy one auditorily. The anticipated contrast in behaviour reinforces the importance of the choice of experimental task in evaluating infant abilities which can span a number of perceptual levels of representation. For various mundane reasons we did not pursue additional bimodal infant speech perception experiments at the time.

I believe that Papillon's model now provides a formal way of understanding this curious finding of /wi∼ju/ because the model allows for representations which are *underspecified in time*, a possibility that is difficult or even impossible to achieve with other phonological theories. That is, this is a representational advance akin to those achieved in Goldsmith (1976) and Raimy (2000) which broadened and refined our notions of temporal relations in phonology.

## 4. A BRIEF ANALYSIS

I propose that the EFP model provides a succinct formal account for these puzzling results, finding an equivalence for the infants in /wi/ ∼ /ju/, and successfully encodes

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

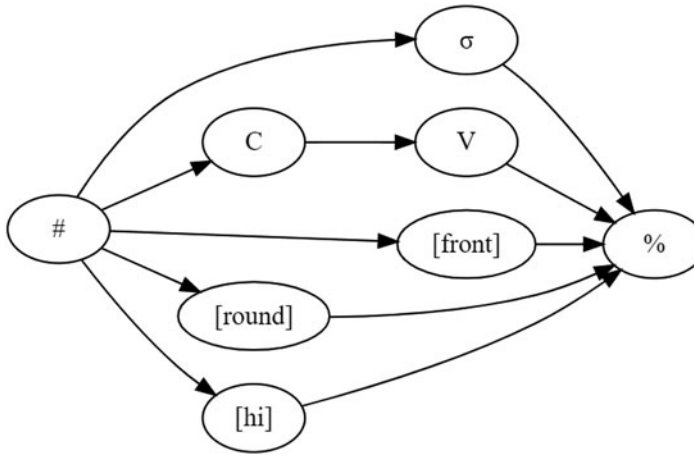


**Figure 5:** EFP representation of an ambiguous /wi∼ju/ syllable.

the informal understanding outlined above. In the EFP framework, we can capture an ambiguous /wi∼ju/ syllable schematically (suppressing irrelevant details) as in Figure 5.

The relative ordering of [front] and [round] is underspecified at this point, as is the temporal extent of the events, but their temporal relationship is not altogether unknown, for they occur in the same syllable. That is, [front] and [round] are not just floating as in the analysis of floating tones in Autosegmental Phonology; their temporal representation is only partially under-determined, perhaps more akin to Firthian prosodies covering an entire syllable. The ultimate ability to discriminate between /wi/ and /ju/ amounts to learning and incorporating the relative ordering of [front] and [round]. When [round] < [front], that is the developing representation for /wi/; when [front] < [round], that is the developing representation for /ju/. Acquiring this kind of serial order knowledge between different features might be fairly difficult, as it is likely that [front] and [round] are initially segregated into different auditory streams (Bregman 1990), and order perception across streams is worse than that within streams. Moreover, it is a contingent fact of English phonology that the vowel system does not allow a [front, round] combination, so we expect some differences in the developmental trajectory of infants learning French or German where the [front, round] combination is licit. One possibility is that the learner would be driven to look for additional temporal relations when the temporally underspecified representations incorrectly predict such cases of "homophony", that is, when they are presented with evidence that /wi/ ≠ /ju/ and therefore conclude that distinct phonological representations are required. If we pursue this idea generally, the EFP graphs will gradually become somewhat more segment-oriented as additional precedence relations are resolved, economizing on the number of events, and leading to later representations, as in Figure 6 for /ju/, where some events have been fused into feature bundles such that [front] < [round] holds.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

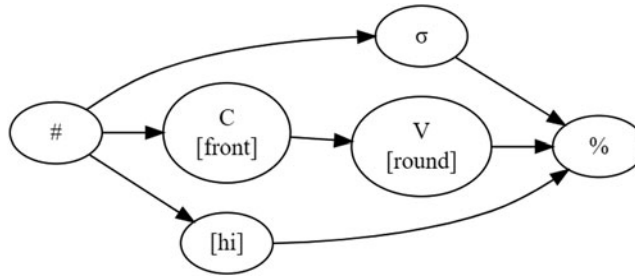


**Figure 6:** EFP representation for /ju/ after feature bundling.

So, the general proposal made here is that the developing knowledge of the relative order of features is initially poor and underspecified for order between features in different "streams" (Bregman 1990). A more general conclusion is that features are not learned from earlier representations involving complete phonetic segments (Mielke 2008, Odden 2019) but rather that features are gradually combined during development to form more nearly segment-sized units. That is, rather than starting with whole segments as the basic entities and then splitting them up into features, instead events with individual features loosely coupled with "longer" events such as syllables (Poeppel 2003) are the starting point, and segments can be constructed by combining events together, as in the fusion of C with [front] in Figure 6. Ultimately this is an empirical question, but the view developed here suggests that temporal precision in speech perception should improve during language development.

## References

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5(2): 183–207.

Avery, Peter, and William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In *Distinctive feature theory,* ed. T. Alan Hall, 41–70. Berlin: Mouton de Gruyter.

Baddeley, Alan. 1986. *Working memory*. Oxford: Oxford University Press.

Baier, Rebecca, William J. Idsardi, and Jeffrey Lidz. 2007. Two-month-olds are sensitive to lip rounding in dynamic and static speech events. In *Proceedings of the International Conference on Auditory-Visual Speech Processing*, L6–2. <https://www.isca-speech.org/archive/avsp_2007/baier07_avsp.html>

Bamberger, Jeanne. 1991. *The mind behind the musical ear*. Cambridge, MA: Harvard University Press.

Barabási, Albert-László. 2016. *Network science*. Cambridge: Cambridge University Press.

van Benthem, Johan F. K. 1983. *The logic of time: A model-theoretic investigation into the varieties of temporal ontology and temporal discourse*. Dordrecht: D. Reidel.

Bertoncini, Josiane, and Jacques Mehler. 1981. Syllables as units in infant speech perception. *Infant Behavior and Development* 4: 247–260.

Bever, Thomas. G., and David Poeppel. 2010. Analysis by synthesis: A (re-) emerging program of research for language and vision. *Biolinguistics* 4(2–3): 174–200.

Bird, Steven, and Ewan Klein. 1990. Phonological events. *Journal of Linguistics* 26(1): 33–56.

Bregman, Albert S. 1990. *Auditory scene analysis*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Browman, Catherine P., and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2): 201–251.

Carson-Berndsen, Julie. 1988. *Time map phonology: Finite state models and event logics in speech recognition.* Berlin: Springer.

Chartrand, Gary, and Ping Zhang. 2012. *A first course in graph theory*. New York: Dover Publications.

Coleman, John, and John Local. 1991. The "No Crossing Constraint" in Autosegmental Phonology. *Linguistics and Philosophy* 14(3): 295–338.

Damian, Markus F., and Jeffrey S. Bowers. 2009. Assessing the role of orthography in speech perception and production: Evidence from picture–word interference tasks. *European Journal of Cognitive Psychology* 21(4): 581–598.

Firth, John R. 1948. Sounds and prosodies. *Transactions of the Philological Society. Philological Society* 47(1): 127–152.

Françoise, Jules, Baptiste Caramiaux, and Frédéric Bevilacqua. 2012. A hierarchical approach for the design of gesture-to-sound mappings. In *Proceedings of the 9th Sound and Music Computing Conference*, ed. Stefania Serafin, 233–240.

Gallistel, Randy. 1980. *The organization of action: A new synthesis*. Hillsdale: Lawrence Erlbaum Associates.

Goranko, Valentine, and Antje Rumberg. 2020. Temporal logic. *The Stanford Encyclopedia of Philosophy.* <https://plato.stanford.edu/archives/sum2020/entries/logic-temporal/>

Goldsmith, John. 1976. *Autosegmental phonology*. Doctoral dissertation, Massachusetts Institute of Technology.

Halle, Morris, and Kenneth N. Stevens. 1963. Speech recognition: A model and a program for research. *IRE Transactions on Information Theory* 8: 155–159.

Heffner, Christopher C., William J. Idsardi, and Rochelle S. Newman. 2019. Constraints on learning disjunctive, unidimensional auditory and phonetic categories. *Attention, Perception, and Psychophysics* 81(4): 958–980.

Idemaru, Kaori, and Lori L. Holt 2014. Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 40(3): 1009–1021.

Idsardi, William J., and Eric Raimy. 2008. Reduplicative economy. In *Rules and constraints in contemporary phonological theory,* ed. Bert Vaux and Andrew Nevins, 149–184. Oxford: Oxford University Press.

Jakobson, Roman, Gunnar Fant, and Morris Halle. 1951. *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Kazanina, Nina, Jeffrey S. Bowers, and William J. Idsardi. 2018. Phonemes: Lexical access and beyond. *Psychonomic Bulletin and Review* 25(2): 560–585.

Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5(2): 275–292.

Kuhl, Patricia K., and Andrew N. Meltzoff. 1982. The bimodal perception of speech in infancy. *Science* 218: 1138–1141.

Lahiri, Aditi, and Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38(1): 44–59.

Lewkowicz, David J. 2004. Perception of serial order in infants. *Developmental Science* 7(2): 175–184.

Li, Chuchu, Min Wang, and William J. Idsardi. 2015. The effect of orthographic form-cuing on the phonological preparation unit in spoken word production. *Memory and Cognition* 43 (4): 563–578.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

Luo, Huan, and David Poeppel. 2012. Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology* 3: 170.

MacNeilage, Peter F. 2008. *The origin of speech*. Oxford: Oxford University Press.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mitterer, Holger, and Eva Reinisch. 2015. Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language* 85: 116–134.

Murphy, Kevin P. 2012. *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Odden, David. 2022. Radical substance free phonology and feature learning. *Journal of Canadian Linguistics* 67(4): 500–551.

O'Seaghdha, Padraig G., Jenn-Yeu Chen, and Train-Min Chen. 2010. Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition* 115(2): 282–302.

Papillon, Maxime. 2020. *Precedence and the lack thereof: Precedence-relation-oriented phonology*. Doctoral dissertation, University of Maryland.

Patterson, Michelle L., and Janet F. Werker. 2003. Two-month-old infants match phonetic information in lips and voice. *Developmental Science* 6(2): 193–198.

Pietroski, Paul. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.

Pietroski, Paul. 2018. *Conjoining meanings: Semantics without truth values*. Oxford: Oxford University Press.

Poeppel, David. 2003. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 41(1): 245–255.

Poeppel, David, William J. Idsardi, and Virginie van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363(1493): 1071–1086.

Poeppel, David, William J. Idsardi. 2011. Recognizing words from speech: The perception-action-memory loop. In *Lexical representation: A multidisciplinary approach,* ed. Gareth Gaskell and Pienie Zwitserlood, 171–196. Berlin: Mouton de Gruyter.

Qu, Qingqing, Markus F. Damian, and Nina Kazanina. 2012. Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences* 109 (35): 14265–14270.

Raimy, Eric. 2000. *The phonology and morphology of reduplication*. Berlin: Mouton de Gruyter.

Rastle, Kathleen, Samantha F. McCormick, Linda Bayliss, and Colin J. Davis. 2011. Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(6): 1588–1594.

Reiss, Charles. 2017. Substance free phonology. In *The Routledge handbook of phonological theory,* ed. S. J. Hannahs and Anna R. K. Bosch, 425–452. London: Routledge.

Sagey, Elizabeth. 1988. On the ill-formedness of crossing association lines. *Linguistic Inquiry* 19(1): 109–118.

Sundara, Megha, Céline Ngon, Katrin Skoruppa, Naomi H. Feldman, Glenda Molina Onario, James L. Morgan, and Sharon Peperkamp. 2018. Young infants' discrimination of subtle phonetic contrasts. *Cognition* 178: 57–66.

Suni, Antti, Juraj Šimko, Daniel Aalto, and Martti Vainio. 2017. Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech and Language* 45: 123–136.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press

Teng, Xianbing, Xing Tian, and David Poeppel. 2016. Testing multi-scale processing in the auditory system. *Scientific Reports* 6: 34390.
Warren, Richard M. 2008. *Auditory perception: An analysis and synthesis*. 3rd ed. Cambridge: Cambridge University Press.

https://doi.org/10.1017/cnj.2022.36 Published online by Cambridge University Press