



ARTICLE

# The blueprint model of production

Scott Nelson <sup>1,2</sup> and Jeffrey Heinz <sup>1</sup>

<sup>1</sup>Department of Linguistics, Stony Brook University, New York, NY, USA

<sup>2</sup>Department of Linguistics, University of Illinois Urbana-Champaign, Urbana, IL, USA

**Corresponding author:** Scott Nelson; Email: [sjnelson@illinois.edu](mailto:sjnelson@illinois.edu)

**Received:** 11 April 2023; **Revised:** 16 October 2023; **Accepted:** 6 April 2024

**Keywords:** phonetics–phonology interface; language production; computational phonology; typed functions; incomplete neutralisation; homophone variation

## Abstract

This article introduces the blueprint model of production (BMP), which characterises the phonetics–phonology interface in terms of typed functions. The standard modular feed-forward view to the interface is that the phonetic form of a lexical item is the output of a phonetic module which takes the output of a phonological module as its input. The central idea of the BMP is that the phonetic form is instead the output of a higher-order phonetics function which takes the phonological function as one of multiple inputs. We explain how understanding the production process this way can account for systematic fine-grained variation in phonetic forms while maintaining a discrete phonological grammar. We present one possible instantiation of the model that simulates incomplete neutralisation, some cases of near-merger, and variation in homophone duration. Consequently, these types of systematic fine-grained phonetic patterns do not necessarily provide evidence against discrete, symbolic phonology.

## Contents

<b>1. Introduction</b>	<b>2</b>
<b>2. The relationship between phonetics and phonology</b>	<b>3</b>
<b>3. The blueprint model of production</b>	<b>6</b>
3.1. Characterising the production process	7
3.2. From assembly line to blueprint: function (de)application	8
3.3. Curryng and uncurrying	10
<b>4. Incomplete neutralisation</b>	<b>12</b>
4.1. Background	12
4.2. How the BMP includes intent	13
4.3. Comparison to the dynamical system approach	15
4.4. Final devoicing in German	16
4.5. Tonal near-merger in Cantonese	19
4.6. Epenthesis in Arabic	21
<b>5. Frequency effects</b>	<b>23</b>
5.1. Background	23
5.2. Adding frequency to the BMP	23
5.3. Homophone duration variation in English	25
<b>6. Conclusion</b>	<b>27</b>
<b>References</b>	<b>29</b>

## 1. Introduction

The division of labour between phonetics and phonology in models of language production is often described such that the phonology handles the discrete and symbolic aspects, while the phonetics transforms the symbols into continuously varying representations relating to some physical dimension. Furthermore, the standard view in generative phonology is what Pierrehumbert (2002) refers to as a ‘modular feed-forward’ architecture. In these types of models, phonetic implementation comes after the phonological grammar and sees only the phonological output.

Certain types of phonetic data, such as incomplete neutralisation (Port *et al.* 1981; Port & O’Dell 1985) and variation in durational properties of homophones (Gahl 2008), are instances where modular feed-forward architectures have struggled to account for the phonetic facts. Notably, it is usually discrete, symbolic phonological grammars that come under attack (Ohala 1990, 1992; Pierrehumbert 2002; Port & Leary 2005). This often results in researchers reconceptualising phonology as continuous, or even eliminating the distinction between the phonetic and phonological modules altogether (Browman & Goldstein 1992; Zsiga 2000; Hayes *et al.* 2004, among others).

We argue for a different approach. Rather than proposing changes to the phonological module directly, we provide an alternate architecture called the blueprint model of production (BMP), which reconceptualises how the modules interact. This reconceptualisation is formalised using typed functions (Pierce 2002), which are related to the lambda calculus (Church 1932, 1933). Under this view, the phonetic and phonological modules are functions.<sup>1</sup> At the core of the BMP is the idea that the phonetic form of a lexical item  $x$  should not be viewed as the output of the composition of a phonetic function  $f$  and a phonological function  $g$  – that is,  $f(g(x))$  – but rather as the output of a higher-order function that takes the phonological function alongside the lexical item as inputs – that is,  $f(g(), x)$ . Furthermore, the phonetic module may take additional inputs  $y$ , such as the speaker’s intent to maintain an underlying contrast, as evidence warrants:  $f(g(), x, y)$ .

Viewing the phonetic module, this way allows for information about both the underlying lexical form and the surface phonological form to be used during the production process. It also allows non-grammatical factors to affect production. With an architecture such as the BMP, phenomena such as incomplete neutralisation and homophone duration variability can in fact be accounted for in a straightforward manner while maintaining a discrete, symbolic phonological grammar. This point is not to say that it has been proven that phonology must be discrete, but rather that the aforementioned evidence does not necessarily imply that phonology must be gradient.

Previous work has also proposed that the phonetic module has access to both the surface representation (SR) and underlying representation (UR) (Goldrick 2000; Gafos & Benus 2006; van Oostendorp 2008; Braver 2019). As §3 explains, the relationship of these proposals to each other and others is made clearer by the typed-function formalism in which the BMP is couched.

Importantly, there are several ways the BMP can be instantiated. This fact means there are distinct levels of analysis, which it is important to be clear about. We reserve the word ‘model’ for higher-level architectures of the phonetics–phonology interface (such as the BMP) and ‘simulation’ for a specific instantiation of a model (such as the ones we present in later sections). This is not standard usage, and many scholars use the word ‘model’ to refer to what we are calling a ‘simulation’. One reason to draw a firm distinction between the two is that it is quite easy to confuse a simulation with the higher-level model, but as McCloskey (1991: 390) warns, ‘any simulation includes theory-irrelevant as well as theory-relevant details; hence, the details of a simulation cannot be identified straightforwardly with the details of the corresponding theory.’ Cooper & Guest (2014) provide a similar warning. Unlike McCloskey (1991), we use the word ‘model’ in place of ‘theory’. This is because theories can exist at different levels (Marr 1982), and across levels, which means a theory may consist of a specific implementation alongside a higher-level architecture.

<sup>1</sup>For the remainder of this article, the terms ‘function’ and ‘module’ are used interchangeably.

Our primary contribution is at the level of the model and not the simulation. Nonetheless, the simulations are important because they illustrate concretely how the model can be implemented. They should not be confused, however, with the model itself. Simulations can be used to test aspects of the model, but simulations come with their own set of assumptions, some of which may be ancillary to the model itself. For example, a simulation may require a parameter whose exact value cannot be derived from the model, and instead is estimated from data deemed relevant. Consequently, critiques of a simulation are not necessarily critiques of the model. It depends on the particulars: a critique of how a parameter's value in a simulation is estimated is not necessarily a critique of the model architecture.

As with the modular feed-forward model, the BMP itself, as an abstract characterisation of the phonetics–phonology interface, has little to say about specific instantiations. Due to this *computational level* description of the BMP (Marr 1982), there are in fact infinitely many possible instantiations. Thus in this sense, the formal model overgenerates. But this is by design: our goal is to describe *capacities* (Cummins 1983; van Rooij & Baggio 2020), not specific implementations. This type of abstract analysis runs into the problem of *multiple realisability* (Putnam 1967; Fodor 1974; Guest & Martin 2023). For example, having the capacity to sort a list of items does not say anything about which of the nearly 50 proposed sorting algorithms<sup>2</sup> is being used. In this same spirit, we are making a claim that language users have a capacity that involves combining lexical information, phonological information and extra-grammatical information when producing speech. While this claim may seem modest, it stands in contrast to the feed-forward model, which prohibits the combination of lexical and phonological information that the BMP provides.

With the computational level description of the BMP in mind, the simulations we present are not without their own specific assumptions. For example, our simulations model variation in the productions of single speakers, not populations of speakers. In addition, they are mostly deterministic because the addition of stochastic variables does not change the overall findings in regards to our specific claim about the model's capacity to account for gradient phenomena with a discrete phonology. In other words, our goal in the simulations is not to find the best quantitative fit of all the variation reported in the literature, but to capture important qualitative attributes sufficient for our argument. This 'proof of concept' is step one in what we envision as a larger research program. In future work, it will be necessary to not only provide a qualitative fit, but a quantitative one as well. Part of this will involve restricting the space of functions that are used when defining specific implementations of the BMP. One area that we think will be especially useful in this regard is computational complexity theory as this has already been proposed as a way to restrict general cognition (Frixione 2001; van Rooij 2008) as well as phonological cognition (Chandlee 2014; Heinz 2018; Lambert *et al.* 2021).

The remainder of the article is laid out as follows: §2 gives an overview of previous accounts of the relationship between phonetics and phonology. §3 provides the formalisation of the BMP using typed functions. The next two sections provide several case studies which show how an instantiation of the BMP with a discrete phonological grammar is able to account for the well documented phonetic properties of incomplete neutralisation and variation in homophone durations. §4 focuses on final devoicing in German (Port & Crawford 1989), tonal near-merger in Cantonese (Yu 2007) and epenthesis in Lebanese Arabic (Gouskova & Hall 2009; Hall 2013). This section further formalises the relationship between incomplete neutralisation and certain cases of near-merger which are shown to be accounted for using the same mechanism. In §5, Gahl's (2008) findings on homophone durations are discussed under the purview of the BMP. The article concludes in §6.

## 2. The relationship between phonetics and phonology

While discussion of the relationship between phonetics and phonology predates *The Sound Pattern of English* (SPE; Chomsky & Halle 1968), SPE is a natural starting point for the current discussion.

<sup>2</sup>[https://en.wikipedia.org/wiki/Sorting\\_algorithm](https://en.wikipedia.org/wiki/Sorting_algorithm)

In *SPE*, it is assumed that the phonology contains rules that map binary features to a scalar value so that the SR of a lexical item is a temporally organised matrix of real numbers corresponding to phonetic features. The phonological grammar therefore contains rules that are both discrete and continuous. It is not explicitly stated whether the two types of rules interact. Additionally, *SPE* assumes that there is a phonetic module that acts as a universal translator, turning the phonetic SR outputs into physical representations.

Keating (1985, 1988) discusses the *SPE* model of speech production further, pointing out that the assumption of a universal phonetics is likely to be incorrect. A main area of focus in her discussion is the tradeoff between enriching the phonological representation with phonetic detail versus having a less phonetically rich SR with language specific phonetic implementation rules. Keating proposes that the grammar contains both phonological and phonetic rules. Kingston & Diehl (1994) argue that speakers use language specific phonetic knowledge to alter their articulations in order to enhance phonological contrasts on the basis of f0 depression around [+voice] segments. This knowledge is implemented outside of the phonological module. Keating (1990) similarly assumes that there are language-specific phonetic rules, but for her, there is phonetic information both inside and outside the phonological module.

It is also possible to consider whether or not we need two separate cognitive modules for phonology and phonetics. A strong argument against separating the two comes in the form of Port & Leary's (2005) article titled 'Against formal phonology'. They argue that a discrete formal symbolic system is unable to account for the variability in phonetic realisation of identical symbols as well as certain temporal contrasts in behavioural data. Since these formal systems cannot simulate the natural language data on their own, Port & Leary (2005) argue against having a formal phonological grammar at all. Ohala (1990) takes a softer approach. He recognises the different types of analysis being done within each domain, but argues that one cannot do phonology without phonetics and one cannot do phonetics without phonology. For him, the two are intertwined, and therefore viewing them as completely separate domains 'is artificial and unnecessarily complicates the study of speech' (Ohala 1990: 156).

Two formal proposals that dissolve the distinction between phonetics and phonology are Flemming's (2001) unified model of phonetics and phonology and Browman & Goldstein's (1992 *et seq.*) theory of Articulatory Phonology (AP). Flemming (2001) develops an Optimality Theoretic (OT; Prince & Smolensky 1993) grammar that operates over scalar phonetic constraints. He argues that phonological assimilation and phonetic coarticulation are essentially the same type of phenomena only with different grain sizes. What is considered to be phonetic coarticulation is just a fine-grained version of the more coarse-grained phonological assimilation (and *vice versa*). The representations in Flemming's model are therefore rich with physical phonetic structure such as formant values (in Hz) and duration (in ms).

AP operates under the assumption that phonetics and phonology are just low- and high-level descriptions of the same dynamical system. At the high level of description, the basic phonological units in AP are gestures. Gestures are task specific goals and therefore defined as the creation of a certain sized constriction in the vocal tract. For example, the word [ta] would be described as a tongue tip gesture that touches the alveolar ridge, a glottal spreading gesture (the default state of the glottis in AP is such that voicing occurs), and a wide tongue body gesture. The tongue tip and glottal gestures would occur in time with one another while the tongue body gesture would be timed to occur after the other two gestures. At the low level of description, each gesture is represented as a second-order dynamical equation and implemented in the task-dynamic model of Saltzman & Munhall (1989). In the task dynamic model, each gesture competes for control of certain articulators while the gesture is active. Since the goal of a gesture is only to create a certain constriction type, the path the articulators take to create a specific constriction are largely dependent on the other gestures simultaneously activated within the dynamical system. From an AP perspective, both phonological and phonetic processes are the lawful consequence of interacting gestures within a dynamical system. While AP is a specific theory that uses dynamical systems, their use more broadly has been successful in describing various interface phenomena (Tuller *et al.* 1994; Gafos 2006; Gafos & Benus 2006; Gafos *et al.* 2014; Roon & Gafos 2016; Łukaszewicz 2021, among others).

If we reject the accounts discussed above and instead favour distinct phonological and phonetic modules, then we are left with deciding where the demarcation point between the two lies. In other words, what exactly is a phonological process and what exactly is a phonetic process? The development of generative phonology coincided with a time when theories of cognition largely involved the manipulation of discrete, symbolic representations (e.g., Newell & Simon 1958). Despite *SPE*'s transformation of features into scalar values, it has largely been assumed that phonological processes are discrete, since the representations are discrete, and that gradience is the result of phonetic processes. This point of view is expressed throughout the literature. For example, Kingston (2019) points to various experimental studies that provide diagnostics for deciding whether a process is phonological or phonetic, all of which involve determining whether or not the process is gradient (Cohn 1993, 2007; Myers 2000; Solé 1992, 1995, 2007).

If gradience is to be the dividing line between phonetics and phonology, there should be a consensus on what type of gradience counts. Gradience has been used in multiple ways when talking about phonology. One way it has been used is in regard to the productivity of phonological generalisations (Albright & Hayes 2006; Ernestus 2011). A second way regards grammatical acceptability judgments (Coleman & Pierrehumbert 1997; Coetzee & Pater 2008). A third way, a focus of this article, is in relation to representations (Smolensky & Goldrick 2016; Lionnet 2017).

Beyond deciding which type of phonological gradience is applicable to the phonetics–phonology interface, Pierrehumbert (1990: 379) points out a logical conundrum for this approach, which is that ‘any continuous variation can be approximated with arbitrary precision by a sufficiently large set of discrete elements’. Consequently, gradience on its own cannot determine whether or not a process is phonetic or phonological.

Gradience notwithstanding, some researchers are perfectly content with interleaving phonetics and phonology. This point of view is represented in the collection *Phonetically Based Phonology* (Hayes *et al.* 2004). The chapters in this book present constraint-based phonological grammars that either are directly inspired by phonetic facts, or, in some cases, directly contain phonetic information. As an example of the latter, Zhang (2004) defines a set of constraints that he calls  $*\text{DUR}(\tau_i)$  that are defined such that for all segments in the rhyme, their cumulative duration in excess of the minimum duration in the prosodic environment in question cannot be  $\tau_i$  or more. He further stipulates that if  $\tau_i > \tau_j$ , then  $*\text{DUR}(\tau_i) \gg *\text{DUR}(\tau_j)$ . The representations therefore must be structured in a way that includes real durational values and not just categorical approximations such as ‘long’ or ‘short’.

In a separate chapter, Gordon (2004) discusses the influence of phonetic properties on phonological syllable weight. Rather than encoding phonetic information directly into the grammar, Gordon shows how phonetic properties of a language could predict weight criteria for tones and syllabic templates. Unlike Zhang's analysis, Gordon retains categorical phonological representations. These examples show that a wide range of views are available when discussing a phonetically based phonology. At one end, there is phonetics *in* phonology, while at the other end, there is something like phonetics *influencing* phonology. Due to this diversity, and unlike Flemming (2001), the essays in this collection are less explicit about the architecture of the grammar, but by using representations and constraints that are phonetic in nature the lines between where phonology ends and phonetics begins are blurred.

In sharp contrast, the substance free phonology framework (Hale & Reiss 2000, 2008; Reiss 2018) demarcates a firm boundary between phonology and phonetics. A core tenet of this framework is that phonological computations should not be based on notions such as phonetic naturalness, typological frequency and markedness. Instead, phonology should be viewed as a symbol manipulator that has one simple goal: to transform the phonological representation according to the rules of the language. For example, maintaining voicing at the end of a phrase has been shown to be difficult due to anatomical reasons (Ohala 1983; Westbury & Keating 1986). A theory of phonology based on notions of markedness or phonetic naturalness would encode this directly into the grammar with a constraint against voiced obstruents in final position. Hale & Reiss (2008: 154–156) argue that this becomes especially problematic if the constraint set is universal, and propose the following thought experiment: imagine in the future, the vocal tract of humans evolves in a way such that it is no longer difficult to



maintain voicing at the end of phrases, but instead it is difficult *not* to maintain voicing at the end of phrases. It would then be phonetically natural to have a process of final voicing, but the grammar already has a universal constraint against final voiced segments because at a previous time they were difficult.

If phonology is completely divorced from such substantive concerns, then one may wonder what connection it has to speech at all. A series of recent articles have clarified that it is only phonological *computations* that are devoid of any substantive influence; phonological *representations* still have phonetic correlates (Volenec & Reiss 2017; Reiss & Volenec 2022). Volenec & Reiss (2017) adopt the fairly standard view that phonological representations are made up of binary feature bundles, but highlight the fact that since phonology is an encapsulated cognitive module (Fodor 1983), its input and output are made up of the same type of representations. Therefore, the UR and SR must both be binary phonological feature bundles. It is only through a separate *transduction* that any type of phonetic representation (PR) can be established. They posit a transducer, which they refer to as ‘Cognitive Phonetics’, which translates the output of phonology (an SR) into a PRs. The PR is ‘is a complex array of neural commands that activate muscles involved in speech production’ (Volenec & Reiss 2017: 270), and feeds the sensorimotor system directly. Furthermore, the Cognitive Phonetics transducer is said to be universal, which recalls *SPE*’s universal translator.

As this section has shown, there are many ways one can think about the interaction of phonetics and phonology. However, not all options have been pursued with the same amount of vigor. We take influence from Gafos & Benus (2006: 924), who write that ‘it is both necessary and promising to do away with the metaphor of precedence between the qualitative phonology and the quantitative phonetics, without losing sight of the essential distinction between the two’. They accomplish this using a constraint-based grammar implemented with dynamical systems.

Rather than commit to a specific implementation, we first provide a more general characterisation of the phonetics–phonology interface based on typed functions (Pierce 2002; Church 1932, 1933). Our general characterisation falls under the Marrian *computational level* category (Marr 1982), as we follow van Rooij & Baggio’s (2020) proposal for adopting a ‘top-down approach’ in modelling psychological capacities. They write ‘Knowing a functional target (“what” a system does) may facilitate the generation of algorithmic- and implementational-level hypotheses (i.e., how the system “works” computing that function)’ (van Rooij & Baggio 2020: 684). Our more specific characterisation used in the simulations is one example of an algorithmic level hypothesis. As previously stated, the simulations are step one in what we envision as a larger research program and are supplementary to the core argument for the structure provided by the BMP.

Crucial to the BMP is conceptualising the phonetic production module as a higher-order function that takes the phonological module as an argument. This does away with ‘the metaphor of precedence’ at the interface while maintaining a distinction between phonology and phonetics (Gafos & Benus 2006). The abstract architecture provided by the BMP allows a diverse range of linguists and researchers in closely related fields working on speech production to interpret their current and future work within this framework. For phonologists specifically, we believe that it provides a way to maintain a simple, discrete phonology that still accounts for gradient production facts. We take this approach when using the BMP in simulations, to show that observed gradience and variation in production does not necessarily imply a gradient phonological grammar. This is ultimately due to the reconceptualisation of how the modules interact within the BMP.

### 3. The blueprint model of production

We will begin this section with a discussion of the production process within generative phonology and then transition into a formal explanation of the BMP. The BMP is best understood as an abstract characterisation of how phonetics and phonology interact during the production process, not unlike how the feed-forward model of production is also an abstract architecture of this interface. As such, there

are many possible ways to *instantiate* the phonetics–phonology interface within the BMP, just as there are many ways to *instantiate* the phonetics–phonology interface within a feed-forward model.

There are two essential points to understanding the BMP. First, it concretely models the production process with multiple simultaneous factors, of which phonology is just one.<sup>3</sup> Second, the whole phonological module is a factor in production, not just the representations it outputs. Like Gafos & Benus (2006), this approach ‘does away with the problematic metaphor of implementation or precedence between phonology and phonetics without losing sight of the essential distinction between the two (qualitative, discrete vs. quantitative, continuous)’. From our perspective, Gafos & Benus (2006) provide one way of accomplishing this. However, it is not the only one possible.

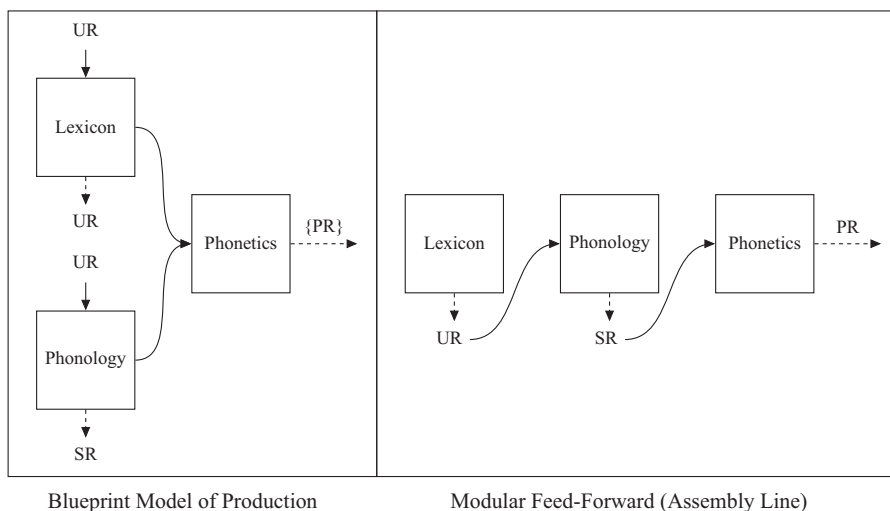
In this context, our contributions are as follows. First, we show how the BMP reconceptualises the relationship between phonology and phonetics. One outcome of this is that the BMP is able to account for gradient phenomena without resorting to a gradient phonology. Consequently, arguments for replacing or removing the phonological module because of systematic phonetic details are not sufficient to displace discrete, symbolic phonology. Second, we are able to relate the BMP to previous work on the phonetics–phonology interface using a type-functional analysis. In particular, we show exactly how the BMP relates to the traditional feed-forward model, as well as earlier proposals which included underlying lexical information in the output of surface forms, which can account for some phonetic effects like incomplete neutralisation.

### 3.1. Characterising the production process

Language production in generative phonology is often assumed to be a modular feed-forward process (Pierrehumbert 2002; Bermúdez-Otero 2007). This type of model is understood as a kind of abstract assembly line: a lexical item is chosen and then is modified through a series of specialised stations until it reaches the end point as a phonetic object that can be pronounced. Since assembly lines are successive, each station is essentially insensitive to the history of the objects it receives. To make this metaphor more concrete, we can imagine that the Lexicon places URs on a conveyor belt which takes them to the Phonology station to be worked on. At the Phonology station, URs are transformed into SRs, and SRs are placed back on the conveyor belt to be taken down the line to the Phonetics station. The Phonetics station receives each SR with no knowledge of its previous history. The role of Phonetics in this instance is to transform each SR into a corresponding phonetic form (e.g., a gradient representation containing acoustic/articulatory instructions). In this example, Phonology acts as an intermediary between the Lexicon and Phonetics. Consequently, when two identical SRs are derived from distinct URs, the Phonetics station must treat those SRs exactly the same way.

Imagine, instead, that Phonology was not a station that a lexical item had to pass through during the production process, but rather a target design that the phonetic module was given alongside a lexical item. In this metaphor, the lexical item is a set of materials, the phonology is a blueprint for what the assembled form should look like, and the Phonetic station is the module which is doing the assembling. The phonology still operates in the same way as in modular feed-forward models: given a UR as an input, it returns an SR as its output. Only now this process does not strictly precede phonetic implementation (cf. Gafos & Benus 2006). This characterisation of the production process situates the phonology in a way that allows it to maintain its primary role of determining the surface form of an UR. It also allows the Phonetics station simultaneous access to both the UR and the phonological instructions on how to modify it. As we explain in more detail later, by ‘phonological instructions’, we simply mean a map from URs to SRs. No other history of a phonological derivation or evaluation is visible to the

<sup>3</sup>The simultaneous, or parallel, view presented here may evoke connectionist models of cognition (Feldman & Ballard 1982; Rumelhart *et al.* 1988; Hinton & Anderson 1989). Our use of simultaneity varies from the connectionist view in that we are talking about it in terms of composing many smaller functions into a larger function. The computation of this larger function does not need to happen in parallel or require a neural architecture. We stress that the functions we propose can be instantiated in any number of ways, including ones which follow connectionist/neural principles and ones that do not.



**Figure 1.** Visual comparison of the architecture for modular feed-forward models and the BMP. Each box represents a function/module. Solid lines represent the inputs to each function while dashed lines represent the outputs of each function.

Phonetics station.<sup>4</sup> The main point here, however, is that under this architecture, the lexical form is not invisible to the Phonetics Station.

Crucial to our analysis is the view that each module can be thought of as a function (Roark & Sproat 2007; Heinz 2018). In the modular feed-forward model, the phonological module is a function that maps a UR to an SR and the phonetic module is a function that maps an SR to a PR. The BMP continues to view the phonological module as a function that maps a UR to an SR, but views the phonetic module as a higher-order function that *takes the phonological module function as an input*. In addition, to generalise over all lexical items, we consider the entire lexicon to be an input to the phonetic module instead of a single UR.<sup>5</sup> The phonetic module is therefore a function with at least two inputs: the lexicon and the phonological module; and one output: a set of PRs  $\{PR\}$ . The two contrasting models are shown in Figure 1. The next section provides a formal definition of the BMP.

### 3.2. From assembly line to blueprint: function (de)application

While giving the phonetic module direct access to the lexicon and phonology may seem like a large departure from the feed-forward model, the BMP can be related directly to the feed-forward model via function application. We also show that the BMP is an abstraction of feed-forward models under the constraint that the representation output by the phonological module *includes* the input to the phonological module (Prince & Smolensky 1993; Goldrick 2000; Revithiadou 2008; van Oostendorp 2008). Our analyses rely on the function type each module computes. Our notation follows from Pierce (2002) which derives from the lambda calculus (Church 1932, 1933). Therefore, we begin with a basic introduction to functions and function types.

<sup>4</sup>The only way around this would be to encode such information into the SR itself. For example, in OT with candidate chains (OT-CC; McCarthy 2007), chains of successive modifications to a form are evaluated. The history could be encoded in the output of the phonology if the whole chain were output instead of just the last representation in the chain. We also later discuss work in which the UR is encoded in some way within the SR (van Oostendorp 2008 and others).

<sup>5</sup>Treating the lexicon as a unitary object is common in computational treatments of morpho-phonology, where the lexicon is represented with a single finite-state transducer (Roark & Sproat 2007; Gorman & Sproat 2021).



A function maps one or more elements in a set  $A$  to elements in a set  $B$  such that each  $a$  in  $A$  maps to at most one element in  $b$  in  $B$ . For a function  $f$  that maps elements from set  $A$  to set  $B$  we write  $f :: A \rightarrow B$ . The phonology function above (or  $P$  for short) would therefore be written as  $P :: UR \rightarrow SR$ . In prose this means ‘the phonology function  $P$  maps URs to SRs’. Note the phonology function  $P$  is agnostic as to the particulars of the representations of  $UR$  and  $SR$ . For example, they could be continuous, discrete, or some combination.  $P :: UR \rightarrow SR$  simply means that the phonological module takes a UR-type thing and returns an SR-type thing.

Functions with more than one argument are written similarly. Addition can be thought of as a function with two arguments:  $add(x)(y) = x + y$ . Its function type would then be written as:  $add :: \mathbb{R} \rightarrow \mathbb{R} \rightarrow \mathbb{R}$ . When reading function types with multiple arguments, everything to the left of the rightmost arrow is an argument and everything to the right of the rightmost (non-bracketed) arrow is the output. The function type of  $add$  can therefore be understood as a map from two real numbers to a single real number. In §3.3, we will discuss how this notation relates to the common  $add(x, y) = x + y$  notation.

Our analysis below relies on two other concepts: higher-order functions and the notion of function application. Functions like the ones described above are first-order functions. These are contrasted with higher-order functions. A higher-order function is a function that either takes as an input another function or returns a function as its output. An example of a higher-order function that takes a function as part of its input is the *map* function.

Given two inputs  $f$  and  $\vec{x}$ , where  $f$  is a function of type  $f :: X \rightarrow Y$  that takes things of type  $X$  as its input, and  $\vec{x}$  is an array of length  $n$  that contains  $x$ ’s  $[x_1, \dots, x_n]$ ,  $map(f)(\vec{x})$  applies function  $f$  to every individual element of  $x \in \vec{x}$  and returns the array  $[f(x_1), \dots, f(x_n)]$ . To give a concrete example, consider the function  $add1(x) = x + 1$  and the array of integers  $[-23, 1, 9, 307]$ . If we were to provide both of these as the input to the *map* function, we would end up with  $map(add1, [-23, 1, 9, 307]) = [-22, 2, 10, 308]$ . The *map* function is not limited to numerical data types/functions and works just as well over strings. For example, for all strings  $w$ , let  $redup(w) = ww$ . Then  $map(redup, [a, ba, cab]) = [aa, baba, cabcab]$ . To summarise, the function type of *map* is given by  $map :: (X \rightarrow Y) \rightarrow [X] \rightarrow [Y]$ .

We now move to a discussion of *function application*. Function application is the act of applying a specific function to an argument, but it can also be thought of as a higher-order function itself. The two arguments for function application would be one of type  $X$  and the other of type  $X \rightarrow Y$  (i.e., a function that maps  $X$  type things to  $Y$  type things). Given these two arguments it would output something of type  $Y$ . For the overall type we would therefore write *function-application*  $:: X \rightarrow (X \rightarrow Y) \rightarrow Y$ . The notion of function application is important for our analysis because it allows us to relate the BMP to the modular feed-forward model.

We now apply these ideas to architectures of language production. Throughout the remainder of this section the following abbreviations are used:  $L$ ,  $P$  and  $A$  as functions representing the **L**exicon, **P**honology and **A**nd **P**honetics (Articulation or Acoustics);  $UR$ ,  $SR$  and  $PR$  to represent **U**nderlying **R**epresentations, **S**urface **R**epresentations, and **P**honetic **R**epresentations. The proposed types are listed in Table 1.<sup>6</sup>

This paragraph describes the steps that turn the modular feed-forward model into the BMP. To start, the phonetic module in the modular feed-forward model has the following type:

$$(1) \quad A_{\text{MFF}} :: SR \rightarrow PR.$$

This idealises the phonetic module as a map from SRs to PRs. Given a UR, the phonology  $P$ , and the definition of function application from above, one can decompose  $SR$  into  $UR \rightarrow (UR \rightarrow SR)$ :

$$(2) \quad A :: UR \rightarrow (UR \rightarrow SR) \rightarrow PR.$$

Next,  $(UR \rightarrow SR)$  is just another way of representing the phonological module:

$$(3) \quad A :: UR \rightarrow P \rightarrow PR.$$

<sup>6</sup>In Figure 1, the Lexicon has type  $UR \rightarrow UR$ . In this case, it can be thought of as the identity function. This is an abstraction to facilitate the analysis.

*Table 1. Types*

Name	Meaning	Type
<i>L</i>	Lexicon	$UR \rightarrow UR$
<i>P</i>	Phonology	$UR \rightarrow SR$
$A_{\text{MFF}}$	Phonetics <sub>MFF</sub>	$SR \rightarrow PR$
$A_{\text{BP}}$	Phonetics <sub>BP</sub>	$L \rightarrow P \rightarrow \{PR\}$
<i>UR</i>	Underlying representation	<i>UR</i>
<i>SR</i>	Surface representation	<i>SR</i>
<i>PR</i>	Phonetic representation	<i>PR</i>

To complete this reconceptualisation, we change *UR* to *L* in order to generalise over the entire lexicon. By doing so, the output is now a set of PRs rather than a single specific representation. This gives us a new type for the phonetics function:

(4)  $A_{\text{BP}} :: L \rightarrow P \rightarrow \{PR\}$ .

The phonetic module is therefore a higher-order function with two arguments: the lexicon and the entire phonological module (a function). As is the case in the modular feed-forward model, the phonology still maps an underlying form to a surface form. Additionally, in both the BMP and the modular feed-forward model an underlying form is ultimately transformed into a PR. The main difference is that the phonology is no longer an intermediary between the lexical form and the phonetic module. Instead, the phonology and the lexical form are both input to the phonetic module.

If it is not clear yet, why we call this model the Blueprint Model, consider this. For every *n*-ary function, there is an equivalent (*n* + 1)-ary relation. Since phonology is a unary function (i.e., it has one input, a *UR*), it can also be envisioned as a binary relation consisting of *UR* and *SR* pairs  $\langle UR, SR \rangle$ . This latter perspective highlights the fact that we can view phonology not as a module that directly shapes the phonetic output, but instead as a set of instructions that informs the phonetic module as to how a given lexical item should be pronounced. In other words, in the same way, one would query a blueprint, the phonetic module queries the phonology as to how a *UR* should be pronounced.

The derivation shown above does not exhaustively represent all the factors that determine production. It simply shows how the BMP relates to the feed-forward model of production. Many other factors have been argued to influence speech production. For example, in the case of incomplete neutralisation it has been argued that the phonetic output is not only a blend of the phonological output (*SR*) and the lexical input (*UR*), but also that this blend can be scaled by extralinguistic factors relating to contrastive intent (Port & Crawford 1989; Ernestus & Baayen 2003; Gafos & Benus 2006). This is an additional factor necessary to adequately account for production. As will be discussed in more detail in §4.2, this is accomplished with the BMP by adding the intent (*I*) as one of the arguments to production:  $A :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$ .

3.3. *Currying and uncurrying*

This section relates the BMP to earlier theories of phonology in which the outputs of phonology include its inputs (Prince & Smolensky 1993; Goldrick 2000; Revithiadou 2008; van Oostendorp 2008). This is precisely the claim made in the original formulation of Optimality Theory, where every element of the phonological input representation is contained in the output (Prince & Smolensky 1993). Under the feed-forward model, the principle of containment ensures that the phonetic module has access to the lexical form, because it can recover it from the output of the phonology. It follows that if the phonological module obeys the principle of containment then the phonetic module is able

to, for example, distinguish between *faithful* word-final voiceless obstruents and *derived* ones (van Oostendorp 2008).

Note that the principle of containment is independent of Optimality Theory *per se*. For instance, it is not difficult to imagine a rule-based theory in which the output of a rule system is a SR presented alongside the UR, which is carried through the derivation. In other words, this principle effectively ensures that the phonological module has something like the type  $P' :: UR \rightarrow (UR, SR)$ , regardless of whether the phonological module is instantiated by a constraint-based grammar, a rule-based grammar, or some other form of grammar.

Strictly speaking, containment theory, and variants such as turbidity theory (Goldrick 2000), do not represent the outputs of phonology as a SR paired with an UR. Instead, the output of a word-final devoicing process for the lexical item /gruz/ would be something more like the sequence [(g,g),(r,r),(u,u),(z,s)]. However, our point is that the UR is recoverable from this representation.

What this means from the perspective of the type-functional analysis is that the containment theory of phonology is an *uncurried* version of the BMP. To explain, consider the fact that since functions in general can return functions, functions with multiple arguments do not need to be given all the arguments at once. If fewer than the totality of arguments is given, then a *function* is returned.

Consider again addition, which we gave the type:  $add :: \mathbb{R} \rightarrow \mathbb{R} \rightarrow \mathbb{R}$ . This can be thought of as the curried version of  $add' :: (\mathbb{R}, \mathbb{R}) \rightarrow \mathbb{R}$ . Whereas *add* takes two arguments, *add'* takes a single argument which is a pair of real numbers. It is always possible to convert between a function which takes one input as a pair of arguments and a higher-order function which takes multiple arguments. This conversion is called *currying* (after Curry 1980). Currying itself can be thought of as a higher-order function, which takes an uncurried function like *add'* and returns the curried version like *add*. The type signature of currying is thus  $curry :: ((A, B) \rightarrow C) \rightarrow (A \rightarrow B \rightarrow C)$ . The argument of the *curry* function is a function mapping (*a*, *b*) pairs to *c*-type things. The output of the *curry* function is a function that takes two separate inputs *a* and *b* and outputs *c*. Consequently,  $curry(add') = add$ , and thus for all *a*, *b*,  $add(a, b) = add'(a, b) = a + b$ .

As mentioned, containment theories of phonology essentially have the type  $P' :: UR \rightarrow (UR, SR)$ . Under the feed-forward model, this output is given to the phonetic module to produce the articulatory representation. Consequently, the phonetic module would have type  $A' :: (UR, SR) \rightarrow PR$ . This is essentially the *uncurried* version of the BMP. Currying *A'* yields a phonetic function of the form in (5):

$$(5) \quad curry(A') :: UR \rightarrow SR \rightarrow PR.$$

Since  $UR \rightarrow SR$  is the function the phonological module computes, (5) can be rewritten as (6):

$$(6) \quad curry(A') :: P \rightarrow PR.$$

Combining (6) and (3) reveals that the BMP can be characterised as shown below:

$$(7) \quad A :: UR \rightarrow P \rightarrow PR.$$

Generalising over the lexicon again, we get the same type for the BMP:

$$(8) \quad A_{\text{BPM}} :: L \rightarrow P \rightarrow \{PR\}.$$

This shows precisely the relation between containment theories of phonology under the feed-forward model and *any* theory of phonology computing functions  $UR \rightarrow SR$  with the BMP. It also highlights the essential difference between the BMP and the modular feed-forward model: the latter serialises phonology and phonetics while the former does not.

The next two sections discuss two empirical phenomena that have been argued to be problematic for theories of language production based on discrete generative models of phonology: incomplete neutralisation (Port *et al.* 1981; Port & O'Dell 1985) and homophone durational variation (Gahl 2008). We argue that these phenomena are not counterarguments to discrete phonological knowledge under the BMP approach to the phonetics–phonology interface. This is due to the fact that the structure of

the interface is itself an analytical assumption that needs to be carefully weighed when discussing the interaction of grammatical and extra-grammatical information in language production.

As many philosophers of science have pointed out, refutation of a given scientific theory is dependent on auxiliary assumptions and shared background knowledge (Quine 1951; Duhem 1954; Popper 1959; Feyerabend 1965; Lakatos 1970). Therefore, phonetic evidence alone does not bear on the nature of phonological knowledge, but rather must be evaluated in tandem with a theory of how phonological knowledge is physically manifested. In other words, phenomena like incomplete neutralisation and variation in homophone duration falsify discrete phonological knowledge only if we assume that the modular feed-forward structure of the interface is a shared assumption (or shared ‘interpretative theory’ in terms of Lakatos 1970). In this way, our analysis aims to show that arguments for gradient phonological knowledge depend on a certain structure of the production function, but there are alternative ways to structure this function that do not require gradient phonological knowledge to account for the same phonetic facts.

#### 4. Incomplete neutralisation

This section first provides background on incomplete neutralisation. After the empirical facts have been laid out, we discuss how the BMP is able to account for the phenomenon by providing one possible instantiation. The section concludes by examining three specific phenomena: final devoicing in German (Port & Crawford 1989), tonal merger in Cantonese (Yu 2007) and vowel epenthesis in Lebanese Arabic (Gouskova & Hall 2009; Hall 2013).

##### 4.1. Background

Final devoicing is probably the best-studied example of a phonological neutralisation process. This is a phenomenon where, at the end of some domain (often syllable or word), an obstruent loses its voicing feature and surfaces as a voiceless segment.<sup>7</sup> It has been attested in a variety of languages including, but not limited to, German (Bloomfield 1933), Polish (Gussmann 2007), Catalan (Wheeler 2005), Russian (Coats & Harshenin 1971) and Turkish (Kopkalli 1994). The data in (9) provide an example from German (Dinnsen & Garcia-Zamor 1971).

- |     |                                  |                               |
|-----|----------------------------------|-------------------------------|
| (9) | a. /bad+ən/ → [badən] ‘to bathe’ | c. /bat+ən/ → [batən] ‘asked’ |
|     | b. /bad/ → [bat] ‘bath’          | d. /bat/ → [bat] ‘ask’        |

In the 1980s, it was discovered that German speakers could discriminate between an underlying voiceless segment and a derived voiceless segment at a rate of 60–70%; further acoustic studies showed that these two types of segments systematically varied along certain acoustic dimensions (Port *et al.* 1981; Port & O’Dell 1985). Acoustically, it was found that the preceding vowel was shorter for underlying voiceless segments, the duration of aspiration noise was longer for underlying voiceless segments, and the amount of voicing into stop closure was longer for underlying voiced segments. These properties make it appear as if the surface form maintained some of the properties of the underlying form. Because the values for the derived voiceless segments were intermediate between a surface voiceless segment derived from underlying voiceless segment and a surface voiced segment in non-coda position, this phenomenon was termed ‘incomplete neutralisation’.

Final devoicing has been extensively studied, and found to be incomplete in many other languages, such as Catalan (Dinnsen & Charles-Luce 1984), Dutch (Warner *et al.* 2004), Polish (Ślwiaczek & Dinnsen 1985), Russian (Dmitrieva *et al.* 2010) and Afrikaans (van Rooy *et al.* 2003). Many other

<sup>7</sup>In this section, we assume a binary [voice] feature but recognise that more specific laryngeal representations have been proposed (Halle & Stevens 1971; Iverson & Salmons 1995; Avery & Idsardi 2001).

processes, such as coda aspiration in Andalusian Spanish (Gerfen 2002), French schwa deletion (Fougeron & Steriade 1997) and Japanese monomoraic lengthening (Braver & Kawahara 2016), have also been found to be incomplete. Strycharczuk (2019) provides a recent review of findings and discusses various hypotheses for the sources of incompleteness.

Returning to final devoicing, Port & Crawford (1989) find that listeners appear to have control over the level of incompleteness of the neutralisation based on communicative context and how salient a contrast is made. In their experiment, they used five different contexts (based on four sentence conditions) to evaluate how the level of neutralisation changed depending on speakers' awareness of the task. Conditions 1A and 1B used disguised sentences where the target word was embedded within a sentence. The 1A task involved participants reading the sentence from a written example. The 1B task used the same sentences, but this time participants were read the sentence and asked to repeat it back out loud to the experimenter. Condition 2 used contrastive sentences where both target words were in the same sentence, but clarifying information was included to differentiate the words. Condition 3 also used contrastive sentences, but removed the clarifying information. In Condition 4, the words were in isolation.

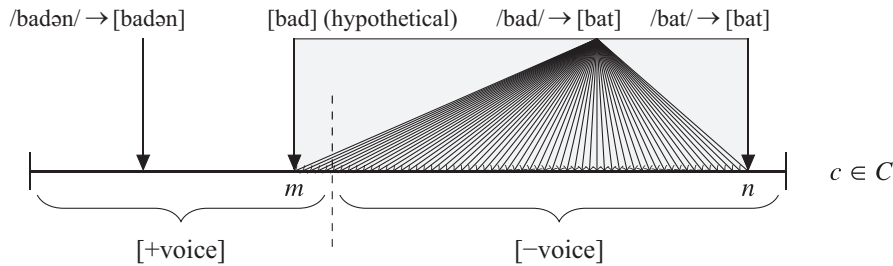
They found incomplete neutralisation in every condition when analysing aggregated speaker data. No difference in the amount of incomplete neutralisation was detected between Conditions 1A and 1B in contrast to previous experiments (Jassem & Richter 1989). In all other cases reported in Port & Crawford (1989), the level of incompleteness increased when the task highlighted the contrast between the two target words. Condition 2 was more incomplete than Conditions 1A and 1B, and Condition 3 was even more incomplete than Condition 2. This makes sense because Condition 2 highlights the contrast, but includes extra material that can aid in distinguishing between the two words. Therefore, speakers may attempt to highlight the contrast with the amount of 'voicing'. Condition 3 meanwhile highlights the contrast, but provides no additional information. In this condition, speakers must use the amount of 'voicing' to make the contrast is salient. Condition 4 also showed a greater degree of incompleteness than Condition 1A/B and was slightly lower than Condition 2.

These data support the idea that speakers have some level of control over how neutralised a segment is depending on the contrastive condition. The pragmatic conditions therefore influence a speakers intent on maintaining an underlying contrast. In their non-linear dynamic approach to production, Gafos & Benus (2006) include a variable called *intent* to account for this fact. For the remainder of this article, we will also use the term *intent* as a cover term indicating pragmatic condition/desire to maintain an underlying contrast.

## 4.2. How the BMP includes intent

§3.2 provided a formal characterisation of the production process. The focus in that section was to show how the BMP conceptualises the phonetic module as taking lexical forms directly alongside the phonology. This means that both UR and SR information is available, something that will be useful in accounting for incomplete neutralisation. That being said, its formulation so far lacks explicit parameters for controlling extra-grammatical factors such as the speaker's intent. The BMP can be updated to include an intent variable in the input, which will scale the production in some way between the UR and SR targets. We use  $I$  for the intent variable, updating the function to:  $A_{BP} :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$ .

In other words, the inputs to the phonetic module reflect multiple factors in production: the lexical form, the phonological instructions, and the pragmatic context. This is a high-level description, and in principle one can see how the phonetic module can account for incomplete neutralisation with this kind of architecture. Nonetheless, it is beneficial to provide one possible concrete instantiation to show how the BMP can simulate the gradient incomplete neutralisation data while maintaining a discrete phonology. The one outlined below is used in the simulations in the remainder of the article.



**Figure 2.** *Hypothetical cue space.*

Recall that the acoustic cues in incompletely neutralised segments are usually in the direction of what might be expected for a phonetic token of the underlying segmental quality. For example, Warner *et al.* (2004) found that Dutch words containing an underlying voiced stop that was devoiced in word-final position were pronounced with a longer preceding vowel than similar Dutch words contained underlying voiceless stops in the same position. Directionality of incompleteness is therefore essential to any account of incomplete neutralisation. Additionally, Port & Crawford (1989) showed that the level of incompleteness seems to be scaled according to pragmatic context. Finally, it is a subset of cues that are found to be incomplete when taking the acoustic measurements. Deciding which cues show up as incomplete and why it is only a subset of cues lies beyond the scope of this article. In subsequent discussion, we talk about a single abstract cue along a one-dimensional space for an individual speaker for expositional simplicity and not epistemic commitment.

Returning to the German final devoicing example in (9), consider a one-dimensional space for some cue  $c$  in the set of all cues  $C$  that signify the voicing contrast for an individual speaker. Imagine dividing the space in such a way that there is a point where every value equal to or less than that point signifies a [+voice] sound while everything greater than that point signifies a [-voice] sound. Within the [+voice] subsection, there may even be different cue values depending on the position of the voiced sound. For example, an intervocalic voiced obstruent may be further away from a specific cue's boundary than a word-final voiced obstruent. It is also the case that the [-voice] subsection can be full of different realisations. In the case of final devoicing, a faithful [-voice] sound may have value  $n$  in the cue space. Likewise, a [+voice] obstruent in final position may have value  $m$  in the cue space.<sup>8</sup>

Since the BMP has access to both UR and SR information, the phonetic form is a blend of the phonetic form given the UR and the phonetic form given the SR. This means that the two points  $m$  and  $n$  provide a theoretical bound on the cue value for the devoiced obstruents in final position. If we assume that the intent variable introduced above controls how much influence the UR or SR has, then the cue  $c$  can in theory surface as any value between  $m$  and  $n$ . Of course, this also depends on the specific implementation of the intent value and scaling process. The next paragraph discusses one way the scaling procedure may be implemented. Figure 2 provides a visualisation of the cue space for the words in (9). Arrows point to possible realisations. Notice that it is only in the alternating case that multiple options exist for a given form.

The main idea sketched above is that the phonetic form is some combination of UR and SR influence. How much influence is given to each is controlled by the intent variable. This is the  $I$  in the  $A_{BP}$  function shown at the beginning of this section. Since intent can be thought of as the percentage that a speaker wants to maintain the underlying form, one way to formalise this notion is as a value in the unit interval  $[0, 1]$ . Here, the lower bound, 0, represents a speaker with 0% intent to maintain the underlying contrast, while the upper bound, 1, represents a speaker who wants to maintain the underlying contrast 100%.

<sup>8</sup>We are assuming here that the phonetic module is able to map a [+voice] sound at the end of a word onto some PR. Since the translation is feature-based, this should not be a problem. The fact that a speaker of a language with final devoicing may never produce a [+voice] sound in this position is due to the phonology and not the phonetics. Anecdotaly, speakers of languages with final devoicing can produce a word-final obstruent as voiced if absolutely forced to do so.



The exact value for cue  $c$  is computed by simply taking a weighted sum of  $c_{UR}$  and  $c_{SR}$ . In Figure 2,  $c_{UR} = m$  and  $c_{SR} = n$ .

One simple way to combine the two values is to use the intent value directly as a weight. This suggests that the scaling process is linear. Another option is to allow for an exponential scaling process. Since incomplete neutralisation typically results in subtle phonetic differences, a linear weighting might indicate that we would expect to see more intermediate cue values when measuring phonetic forms. Exponential scaling still allows for the UR value to have an effect on the phonetic form, but only in circumstances where there is a high intent value will it result in anything other than subtle variation. The following formula provides an exact formulation of exponential scaling where  $\alpha > 0$ :

$$(10) \quad c = c_{UR} \times I^\alpha + c_{SR} \times (1 - I^\alpha).$$

This formula has desirable properties. First, when  $I = 0$ , there is no effect of the UR on the output, and when  $I = 1$  there is no effect of SR on the output. While this may seem trivial, it does match the informal explanation of intent. Second, since the scaling weights sum to 1 it is impossible for  $c$  to fall outside the bounds set by  $c_{UR}$  and  $c_{SR}$ . If we assume  $c_{UR} < c_{SR}$ , then for any arbitrary values of  $I$ , the only way for  $c > c_{SR}$  is to have  $c_{UR} \times (1 - I^\alpha) > c_{SR} \times (1 - I^\alpha)$ . But this reduces to  $c_{UR} > c_{SR}$  which is a contradiction. This proof works the same way to show how it would not be possible to get a value lower than  $c_{UR}$  in this same scenario. Third, because the  $\alpha$  parameter is tied to a specific cue, it provides a potential explanation for *how* only certain cues can be incomplete. Again, we choose not to speculate on *why* certain cues show up as incomplete while others do not, but do provide this mechanism as a way to include the variation.

When  $\alpha = 1$  there is a linear effect of the UR on the final output. In this case, the percent influence of the UR is equal to the intent value. As  $\alpha$  increases, the influence of the UR becomes less and less for lower intent values. For high values of  $\alpha$ , it is only high values of intent that will allow for the UR to have any influence on the output form. This exponential scaling potentially explains why the effects of incomplete neutralisation are subtle, and also that, under extreme circumstances, speakers can produce something very UR-like (see footnote 8).

### 4.3. Comparison to the dynamical system approach

As discussed above, the approach of Gafos & Benus (2006) has been rather influential on our formulation of the BMP. When accounting for final devoicing, they describe a constraint grammar based in nonlinear dynamics that contains separate equations for a markedness constraint (pulling the system towards a voiceless surface form) and a faithfulness constraint (pulling the system towards a voiced underlying form). The two approaches share many aspects: the lexicon and grammar are expressed in terms of functions, extragrammatical information can enter the computation, and there is no direct precedence of phonology over phonetics. Fundamentally, though, these ideas are expressed in two different mathematical frameworks. We use the language of functions and function types as used in programming language theory and other areas of theoretical computer science and discrete math. Gafos & Benus (2006) use the language of nonlinear dynamics, which allows them to simultaneously express discrete and continuous aspects of a complex system.

These two approaches make very different philosophical claims about cognition in terms of the symbolic nature of cognitive knowledge. One large advantage to the dynamical systems approach when it comes to phonetics and phonology is the fact that there is no extra translation mechanism needed to turn symbolic phonological knowledge into continuous phonetic substance. Nonetheless, we believe it is instructive to imagine an instantiation of the BMP which draws directly from dynamics of Gafos & Benus (2006).

Consider an instantiation of the BMP where the type realisations for the URs, the SRs, and the phonetics representations, are the same (i.e.,  $UR = SR = PR$ ). In particular, these representations reference specific phonetic cues which are given by differential equations of the form  $\dot{x} = f(x) = -dV(x)/dx$ , which describe a time-invariant first-order dynamical system in control of a cue, where  $f(x)$

is a force function acting upon the state of the system and  $V(x)$  is the related potential. For concreteness, consider the force function  $\dot{x} = F(x) = x^{REQ} - x$  with corresponding potential  $V(x) = x^2/2 - x^{REQ}x$ , where  $x^{REQ}$  is a set of target values  $\{-x_0, x_0\}$  which are fixed based on the positive and negative values of some binary feature. If used as the functions for *UR* and *SR* in the BMP, they would represent the underlying and surface values of the relevant phonetic cue. The phonetic module in the BMP could then combine them as described in equation (10) to get the final *PR* form. Ultimately, Gafos & Benus (2006) take a different approach to their dynamics. They use a tilted anharmonic oscillator to formalise a markedness force function:  $\dot{x} = M(x) = -k + x - x^3$  with corresponding potential  $V_M(x) = kx - x^2/2 + x^4/4$ . In addition they use a parameter  $\theta$  to express contrastive intent within their faithfulness force function as a way to influence the ‘underlying’ form:  $\dot{x} = F(x) = \theta(x^{REQ} - x)$ ; and corresponding potential  $V_F(x) = \theta x^2/2 - \theta x^{REQ}x$ . They then add the two forces together:  $\dot{x} = M(x) + F(x)$ .

Our point with this exercise is to emphasise clear parallels between the BMP and the specific approach of Gafos & Benus (2006). Where the BMP associates a cue value with a *UR*, Gafos & Benus have a force equation that places a fixed point at the corresponding lexical/underlying value for voicing (faithfulness). Where the BMP associates a cue value with a *SR*, they have a force equation that pulls the system towards a point corresponding to the surface value for voicing (markedness). In both cases, these values/equations are summed, but in the case of dynamical systems the scaling controlled by the contrastive intent happens within these equations themselves and not with an external parameter as is the case for the BMP.

What we continue to stress in this article is that language production involves the interaction of lexical, phonological, and extragrammatical factors which the modular feedforward model fails to capture. Since this idea can be expressed using different types of mathematical formalisms, we believe that this idea is not a property of the specific mathematical implementation, but rather a property of the high-level architecture (a ‘model’ in our terms). Our simulations below stress this fact by showing that a non-dynamical implementation involving a discrete phonological grammar can also account for the qualitative behaviour of individual language users.

In the remaining parts of this section, we present three case studies to show how our instantiation of the BMP can account for the phonetic facts of incomplete neutralisation in three distinct processes: final devoicing in German, tonal processes in Cantonese, and epenthesis in two dialects of Arabic. These three case studies also highlight the relationship between incomplete neutralisation and near-merger. We show that in all cases, the data can emerge from the same system, therefore providing a unified explanation for these phenomenon, despite previous researchers positing different mechanisms.

Since our simulations do not use dynamical systems, they provide an alternative approach to the interface. However, we are not proposing these simulations *in opposition* to the dynamical approach. While we believe that the success of our simulations *sufficiently* captures important qualitative aspects of production, it does not *necessarily* negate the dynamical systems approach to the interface. Our simulations are introduced with the aim to establish that the BMP is a framework with many possible instantiations. This further helps clarify which level of analysis provides the source of explanation for speaker behaviour in the phenomena we study. In our opinion, it is at the level of the model and not the level of the simulation.

#### 4.4. *Final devoicing in German*

The intent argument was added to the BMP in order to account for Port & Crawford’s (1989) results from German that show that the level of incompleteness can vary based on pragmatic factors. This section shows how the intent argument and the  $\alpha$  parameter can interact to simulate their findings. The simulation below focuses on burst duration, which was the main cue Port & Crawford (1989) found to be incomplete, and closure duration, which they found to be complete. The exact cues found to incompletely neutralise have varied across studies. For example, there have been conflicting results about whether or not preceding vowel duration is an incomplete cue in German final devoicing.

**Table 2.** Data from Port & Crawford (1989) for neutralised final stops by condition. Ratio indicates the mean value of /d/ divided by the mean value of /t/.

Condition		Closure duration (Mean)	Ratio	Burst duration (Mean)	Ratio
1A	/d/	54.72	0.91	20.08	0.78
	/t/	59.84		25.59	
1B	/d/	50	0.91	16.54	0.58
	/t/	54.72		28.35	
2	/d/	68.89	1.02	32.87	0.83
	/t/	67.52		39.37	
3	/d/	86.22	1.03	25.20	0.29
	/t/	83.46		85.63	
4	/d/	88.98	0.99	59.06	0.89
	/t/	89.93		66.51	

Nicenboim *et al.* (2018) ran a statistical meta-analysis using a Bayesian random-effects regression model and found a main effect that supported vowel duration as a significant cue of incomplete neutralisation. Port & Crawford (1989), on the other hand, reported preceding vowel duration as being complete in their findings.

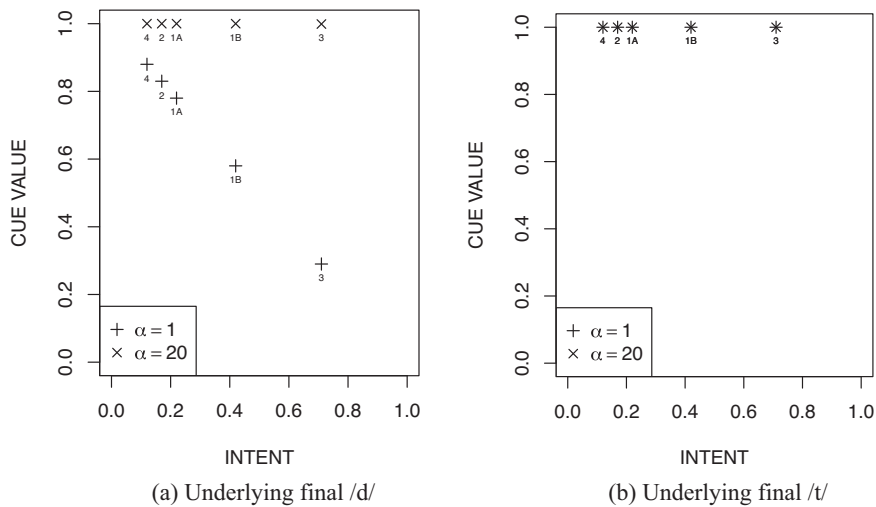
Our assumption, based on the results of Port & Crawford (1989), is that the level of incomplete neutralisation can be dynamically controlled based on pragmatic context. With this in mind, we provide a simulation of their results to highlight the distinction between cues that are complete and cues that are incomplete, while accounting for the pragmatic scaling based on intent. Since the conclusions of Port & Crawford (1989) and the meta-analysis conflict with respect to vowel duration, we avoid this cue altogether. Ultimately, our simulation results do not depend on the specific cues that neutralise incompletely or not, but rather on the working assumption that cues can vary in this way at all.

Mean values for both closure duration and burst duration for each neutralised final stop pair and each condition are shown in Table 2. These data are taken directly from Port & Crawford (1989: 265, Table 1).<sup>9</sup> The ratio columns were added by dividing the final /d/ values by the final /t/ values in each condition. Since only the voiceless target is recoverable from the phonetic data in final position, we rely on the ratio to relate surface final /d/ to some hidden underlying target.

The results are simulated by assuming a single intent value for each pragmatic context, but a different  $\alpha$  value for each cue in the scaling function. Abstracting away from specific values, we assume for all cues that a value of 1 is equal to the voiceless target and a value of 0 is equal to the voiced target. Since our focus is on accounting for the different levels of incompleteness, using ratios abstracts away from the condition specific variation. Therefore, the ratios reported in Table 2 can be used to estimate intent values.

Each subfigure in Figure 3 shows the estimated cue values for both burst duration and closure duration. In general, the ratio of closure duration for underlying /d/ segments to underlying /t/ segments was 0.91 or higher for each condition. Since burst duration (represented as +) was found to vary significantly between derived and faithful surface /t/ segments in the pooled data, but closure duration (represented as ×) was not, the  $\alpha$  parameter was set to 1 for burst duration and 20 for closure duration.

<sup>9</sup>This is data aggregated across multiple speakers. Our simulation treats this as one speaker. We discuss how to lift the simulation to populations at the end of this section.



**Figure 3.** Simulated cue values for Port & Crawford's (1989) results. The left plot shows values for /d/-final words; the right plot shows values for /t/-final words. The symbols + and × represent burst and closure duration, respectively.

The ratios for burst duration varied from 0.29 for condition 3 to 0.88 for condition 4. Intent values were determined by subtracting the burst duration ratios from 1. The resulting plot shows that even with largely varying Intent values, the  $\alpha$  parameter can make it so only a single cue shows up as being incomplete.<sup>10</sup>

From the figures, it is possible to compare both within plots and between plots, resulting in four comparisons. Based on Port & Crawford's (1989) data, we expect variation between the two cues for final /d/ and no variation between the two cues for final /t/. We should also expect to see variation between final /d/ and final /t/ for burst duration, but no variation between final /d/ and final /t/ for closure duration. In Figure 3a, the cue values are shown to vary between burst duration (+) and closure duration (×), as expected. Cue values close to 1 indicate that the final /d/ that has been neutralised has acoustic properties that are basically similar to the faithful final /t/ segments. The closure duration cue values for /d/ are close to 1, as they are for /t/. For /t/, burst duration is close to 1 as well. Again, this is expected given the data. While it may seem trivial that all of the underlying final /t/ values are right at 1, given that they were the denominator for determining ratios, these values were derived with the same formula that derived the final /d/ values. That is, the same  $\alpha$  and same intent values were used, but the formula ensures that final /t/ values are unaltered.

This simulation shows that the interaction of the intent and the alpha parameters captures the aggregate behaviours observed by Port & Crawford (1989), where burst duration was incompletely neutralised and varied according to pragmatic context, while closure duration did not. A reviewer points out that closer inspection of individual behaviour in Port & Crawford (1989) shows that there was variation across individuals in the manifestation of cues in relation to incompleteness as well as interpretation of pragmatic context. Our simulation could be modified to simulate this kind of population-level behaviour in different ways. One way would include a probability distribution over the  $I$  and  $\alpha$  parameters from equation (10). While this may better capture group behaviours, we don't believe that would provide any further insight into what we find important: the capacity of the individual

<sup>10</sup>In their discriminant analysis, Port & Crawford (1989) found that underlying /d/ was more easily recognised in condition 2 than in condition 1. This goes against the acoustic data presented in the article that shows that conditions 1A and 1B were more incomplete based on what the ratios suggest. We thank an anonymous reviewer for pointing out that this is likely due to glottal pulsing not being included as a cue in the discriminant analysis.

speakers. The simulations therefore are deterministic under the assumption that a given speaker, with specific intent values, and specific alpha values would act in a certain way. Likewise, we could add an error term to account for noise in the system not captured by the simulation, but this again would not change the interpretation of the qualitative behaviour the simulation exhibits.

The overall structure of the BMP allows for lexical influence on phonetic form. It also accounts for incomplete neutralisation while maintaining a singular phonological devoicing rule, *contra* Port & Crawford (1989), who claim that their data refutes such a possibility. They write, ‘One can apparently only write accurate rules for German devoicing by making them speaker-dependent and by employing a very large set of articulatory features to capture the detailed dynamic differences between the speakers’ implementation of the contrast’ (Port & Crawford 1989: 280). This interpretation follows from conceptualising the phonetics–phonology interface in terms of the modular feed-forward model, but it does not follow from conceptualising it in terms of the BMP. This is because the BMP is able to capture ‘dynamic differences between the speakers’ implementation of the contrast’ by recognising multiple *simultaneous* factors influencing phonetic production. One factor is the lexical form, and another can be a discrete phonology with a singular devoicing process. Port & Crawford (1989) show that pragmatic context is a necessary ingredient, which is formalised in the BMP as intent. Individual speakers’ implementation of contrast does not need to be encoded in the phonological grammar, because with the BMP speakers have access to the contrast outside of the phonological module. This highlights the roles that both competence and performance play in the production process (cf. Chomsky 1965). In both cases, there is knowledge that is being used during implementation: lexical knowledge, discrete phonological knowledge and a continuous representation of contrastive intent. It follows that under the BMP a continuous phonetic output does not require a continuous phonological grammar.

#### 4.5. Tonal near-merger in Cantonese

The similarity between incomplete neutralisation and near-merger has been well documented (Ramer 1996; Winter & Röttger 2011; Yu 2011; Braver 2019). While the term ‘incomplete neutralisation’ emerged from the phonetic and phonological literature, the term ‘near-merger’ originated in sociolinguistics. Near-merger can be traced back to Labov *et al.* (1972) and their work on New York City English. Words like *source* and *sauce* were reported to be identical by speakers, but then consistently produced with slightly different phonetic forms. The term ‘near-merger’ is therefore usually used when two classes of sounds are perceived as being of the same category, but produced with subtle variation.

One aspect of Port *et al.*’s (1981) argument for incomplete neutralisation was that listeners could correctly guess the specific word at an above-chance level, highlighting the perceptibility of the contrast. This suggests that the primary difference between incomplete neutralisation and near-merger is whether the difference is perceptible. There is also the synchronic versus diachronic distinction. Near-merger has been used by sociolinguists to explain sound change, while incomplete neutralisation is often related to the active production process.

Alternations also help distinguish the two. In the *source–sauce* example, there is no alternation driving the neutralisation, but incomplete neutralisation is dependent on there being an alternation. Regardless of whether or not these two phenomena are one and the same, we believe that certain cases of near-merger can be explained with the same mechanisms we have developed for incomplete neutralisation using the BMP.

Tonal near-merger in Cantonese as discussed by Yu (2007) is one such case. Unlike the *source–sauce* example, it involves morphological alternations called *pinjam*. These alternations involve a non-high-level tone turning into a mid-rising tone.

- (11) a. sou<sup>33</sup> ‘to sweep’ → sou<sup>35</sup> ‘a broom’  
       b. pɔŋ<sup>22</sup> ‘to weigh’ → pɔŋ<sup>35</sup> ‘a scale’  
       c. ts<sup>h</sup>əŋ<sup>11</sup> ‘to hammer’ → ts<sup>h</sup>əŋ<sup>35</sup> ‘a hammer’

The derived mid-rising tones of these *pinjam* words were compared with lexical mid-rising tones in lexical near-minimal pairs. The  $f_0$  value at the onset of the tone, the inflection point, and peak of the rise were all found to be higher for the *pinjam* words. Furthermore, a follow-up study on this phenomenon showed that listeners were unable to tell the two types of mid-rising tones apart, thus giving it its near-merger status.

On first glance, this seems to make the opposite prediction of what might be expected given the UR/SR scaling account we have developed so far. The derived *pinjam* 35 tones should be lower than the lexical mid-rising tones since they (potentially) correspond with a non-high-level tone. A closer look shows that the phonological analysis involves an underlying floating high tone:  $p\alpha ng^{22}(55) \rightarrow p\alpha ng^{35}$  ‘a scale’ where parentheses indicate a floating tone. In this case, it may be interpreted that the reason that the *pinjam* mid-rising tone has higher  $f_0$  values than the lexically specified mid-rising tones is due to the inclusion of an underlying high tone.

Yu (2007) explains the data using an exemplar model with further support coming from contracted syllables (sandhi). The morphemes /tsɔ/ and /təkʰ/ both surface with a mid-rising tone in contracted syllables:

- (12) a.  $p\alpha \eta^{22} \text{ ts}\alpha^{35} \rightarrow p\alpha^{35}$  ‘to weigh (PERF)’  
 b.  $p\alpha \eta^{22} \text{ t}\epsilon k^{55} \rightarrow p\alpha^{35}$  ‘to weigh (POTENTIAL)’

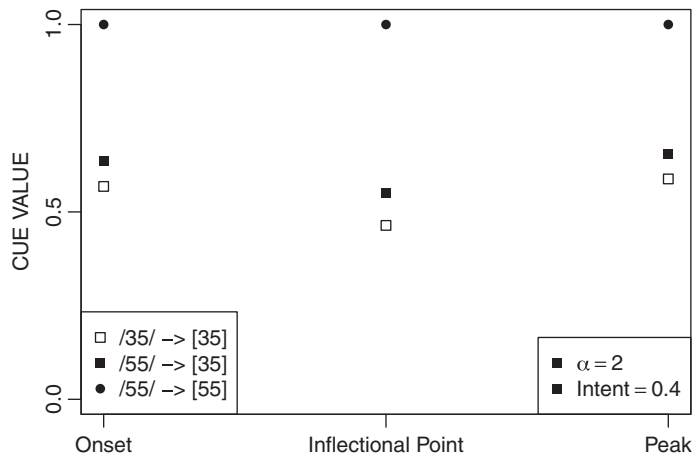
What makes it interesting is that /tsɔ/ has an underlying mid-rising tone while /təkʰ/ has an underlying high tone. The  $f_0$  value at all three points was found to be higher for the mid-rising tone derived from the underlying high tone than for the mid-rising tone that was underlyingly mid-rising. In the BMP, this is exactly what would be expected. That is, a surface mid-rising tone that was derived from an underlying high tone should have its  $f_0$  values raised, given a non-zero intent value. Despite the exemplar interpretation, Yu (2007: 207) recognises this fact and writes, ‘Thus, the extra-high  $f_0$  of the [derived mid-rising tone] can be interpreted as the retention of the tonal profile of an underlying [high] tone.’

Figure 4 shows simulated data for the sandhi process. Our goal with this simulation is to show the qualitative ‘in between-ness’ of the derived mid-rising tone at all three measured points. We take the same approach as in §4.2, where we abstract to a  $[0,1]$  cue space. In this example, 1 corresponds to a high tone (5) and 0 corresponds to a low tone (1). Using an  $\alpha$  value of 2 and Intent value of 0.4, the values for three types of mappings are shown: a faithful mapping of the high tone (/55/→[55]), a faithful mapping of a mid-rising tone (/35/→[35]), and an alternation where an underlying high tone turns into a mid-rising tone (/55/→[35]). Shapes indicate surface tone: squares are mid-rising and circles are high. Color indicates underlying tone: white is mid-rising and black is high. The derived mid-rising tone is therefore represented by a black square.

In the simulation, the faithful mappings are unaffected by the  $\alpha$  and intent values, and the values for the alternation mapping are an interpolation between these two extremes. This shows once again that this instantiation of the BMP captures important qualitative aspects of this tonal phenomenon.

A reviewer points out that our simulation fails to capture the size of the difference at different points. We agree that this is a shortcoming of the specific implementational choices. Nonetheless, our goal was to simulate the in between-ness and not the exact magnitude. One potential fix would be to vary the cue value for [55] at each point. Currently, the size of the difference is based on the size of the difference between the [35] target and the [55] target. It seems reasonable to say that the [55] peak is the true maximum (1) value on the cue dimension, and the onset and inflection point values are lower. Therefore, if we make the [55] peak relatively high enough, the difference at the peak will always be greatest. Since there is no data provided by Yu (2007) on the phonetic properties of the [55] tone, we leave this for future work. We stress once again that this specific implementational choice does not actually impact any claims about the model structure (i.e., the type of information available and the way it combines).





**Figure 4.** BMP prediction for Cantonese tonal merger: Simulated cue values for Yu's (2007) tone-sandhi data.

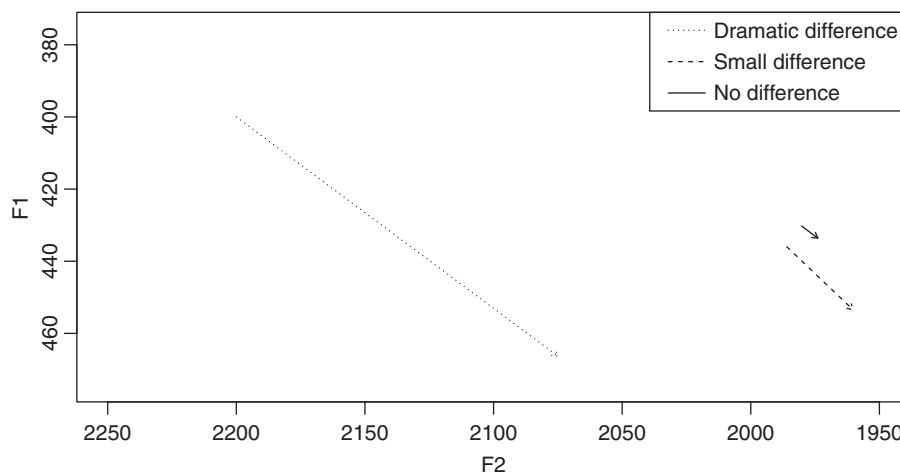
Yu (2007) also found that the mid-rising tone derived from an underlying high tone in the contracted syllables had higher  $f_0$  values than the *pinjam* mid-rising tone (also derived from an underlying high tone). The exemplar model explains this data with an averaging effect. An alternative explanation is that the act of syllable contraction highlights the underlying form more directly than *pinjam*, and therefore speakers are more likely to have a higher intent value, thus pulling the final phonetic form towards the underlying high tone values.

This section shows that while near-merger and incomplete neutralisation have been described as two separate phenomena, they can, in certain cases, emerge from the same basic system. The BMP only relies on a correspondence between underlying and surface forms which is anticipated through the phonological mapping. Any phonological change, whether it be morphologically driven or otherwise, will predict the same type of phonetic effects in this model. The phonetic distribution of any segment should therefore be bounded between what we would expect given the underlying form and the actual surface form.

#### 4.6. Epenthesis in Arabic

Lebanese Arabic speakers epenthesise an [i] vowel to break up word final CC clusters. Gouskova & Hall (2009) performed an acoustic study that had speakers pronounce words with underlying forms /CVCC/ and /CVCiC/. Words of the first form are pronounced the same as the second form due to the epenthesis process. In both cases, the last vowel is an [i]. Measurements of the acoustic properties of these vowels found that the epenthetic [i] showed statistically significant differences in duration and occasionally F2 frequency when compared to [i] tokens that were present in the underlying form. Notably, the authors write, 'epenthesis introduces something *less than an [i]*: the vowel is backer and shorter, all properties that would make this vowel closer to [i] or [ə] – and, arguably, to zero' (emphasis original). While they use Optimality Theory with Candidate Chains (OT-CC; McCarthy 2007) to explain these findings, the fact that the acoustic properties of the epenthetic vowel are more similar to zero is expected given the BMP.

Since the BMP relies on a segmental correspondence between underlying and surface forms, the correspondent of an epenthesised segment is arguably zero. The spatial cues for a zero segment may be the neutral articulatory values for the speaker/language, but the durational cue would be zero. This means that phonologically epenthetic vowels would range from 0 ms when Intent is 1 to the target duration for an [i] vowel when Intent is 0. If the level of Intent is between 0 and 1, then the duration of the epenthetic vowel will always be closer to zero, which is exactly what Gouskova & Hall (2009) find.



**Figure 5.** Simulated formant values for lexical and epenthetic [i] vowels based on Hall (2013) for a dramatic-difference speaker, a small-difference speaker and a no-difference speaker.

Hall (2013) follows up on Gouskova & Hall's (2009) work with a larger number of speakers. In the original study, it was found that the level of incompleteness varied from person to person and this finding was strengthened in the follow up study, most notably in relation to formant values. In fact, no difference in duration was found between the lexical (61 ms) and epenthetic (60 ms) vowels at the group level.<sup>11</sup> Hall (2013) hypothesises that this may be a result of the faster speech rates used in data collection for this study than those used in data collection in Gouskova & Hall (2009). For this reason, our simulation focuses on the formant values.

When comparing the mean value of epenthetic versus lexical [i], Hall (2013) groups speakers into two categories: dramatic difference and non-dramatic difference. She further claims that the non-dramatic difference ranges from speakers with a small difference to those with no difference at all. We therefore use three groups in our simulation: dramatic difference, small difference and no difference. Notably, the dramatic-difference speakers all have a higher/fronter lexical [i] compared to the other speakers. We can take this into account in a simulation by having the dramatic speaker have a different surface [i] target than the other two types of speakers. Figure 5 shows the simulated F1 and F2 values for each type of speaker. The starting point of the arrow is the lexical [i] values and the end point of the arrow is the epenthetic [i] values.

This paragraph lists the parameters used to determine the values in the scaling simulation. For the dramatic-difference speaker, lexical [i] was assigned the  $F1 \times F2$  vector (400, 2200); the other two speakers were each assigned the vector (450, 2000), to indicate a more central vowel. Some noise was added to the second two speakers' vectors to provide visual separation in the plot. This is because otherwise the lines on which each arrow sat would be overlapping. Since there was more movement along the F2 dimension in Hall's (2013) data, the F2 cue was determined with an  $\alpha$  value of 2 while F1 was determined with an  $\alpha$  value of 2.4 (since a higher  $\alpha$  leads to less incompleteness). Finally, Intent levels were set to 0.5, 0.3 and 0.15 for the dramatic-difference, small-difference and no-difference speakers. This is not the only way to simulate the different type of speakers. For example, it is possible to have a single Intent value and instead have the  $\alpha$  levels for different cues vary across speakers. There is not enough empirical data to choose between simulation strategies here. Therefore, we again emphasise that this simulation is only one way to instantiate the BMP.

Another dimension that can affect the simulation results is the spatial parameters of the underlying zero form. This also varies drastically based on what choices are made in regard to PRs. If PRs are

<sup>11</sup> Individual differences were not reported.

acoustic targets, then a zero morpheme would have to have some type of acoustic target even if its duration is also 0. One plausible set of values is that corresponding to the default/neutral segment in the language (Archangeli 1984; Broselow 1984; Pulleyblank 1988; McCarthy & Prince 1994). In the simulation above, we chose a neutral vowel (schwa) as the basis for the F1 and F2 targets, but this is ultimately an implementation choice rather than an architectural choice. Our main point continues to be about the latter, but by being explicit we can investigate consequences of the former. Ultimately, it may make more sense to think about zero morphemes in terms of articulation. A durationless target may still have spatial targets, but they can be thought of as the neutral position of the articulators – which would also lead to the vowel being more central.

In the original study, Gouskova & Hall (2009) claim that the phenomenon at hand is a case of incomplete neutralisation, but Hall (2013) suggests that what is going on is more likely to be near-merger. Regardless of what it should be called, there is some type of intermediary effect between an underlying form and a surface form, and this is what the BMP predicts by having access to the lexicon, the phonological grammar and the pragmatic context in which utterances are being made. The BMP is agnostic to perception, and therefore the perceptibility of a given token plays no role in the synchronic phonetic realisation. This is what allows for a unified explanation of German final devoicing, Cantonese tonal merger and Lebanese Arabic epenthesis.

## 5. Frequency effects

### 5.1. Background

Up to this point, we have discussed scenarios where various lexical items have identical surface forms but phonologically distinct underlying forms. In these cases, the variation between underlying and surface forms allows for interpolation between the two. We now turn our attention to a different scenario: homophony. It has been reported that many homophonic pairs have subtle phonetic differences, most notably along the temporal dimension (Walsh & Parker 1983; Losiewicz 1995; Gahl 2008; Lohmann 2018a,b). Like neutralised pairs, homophones share the same surface phonological form, but unlike neutralised pairs there is no guarantee that they have diverging underlying forms. Nonetheless, the architecture of the BMP offers an explanation for the phonetic variation of homophones.

Frequency has long been known to play a role in the phonetic realisation of phonological units (Fosler-Lussier & Morgan 1998; Bybee 2001; Jurafsky *et al.* 2001; Bell *et al.* 2009). Leslau (1969) reports that the Arab grammarians were attuned to this phenomenon as they noted that more frequent words become ‘weaker’. Another dimension that can play a role in this phenomenon is part of speech. For example, words like *road* (N) and *rode* (V) have been found to vary in their pronunciation (Bell *et al.* 2009). Gahl (2008) looked at non-function word homophone pairs such as *time* (N) and *thyme* (N) and found that there was a difference in duration that correlated with frequency of the lemma. This clearly implicates lexical frequencies in production. Based on these findings, Gahl (2008) rejects discrete, symbolic lexical representations and instead argues for an exemplar-based organisation of the grammar.

### 5.2. Adding frequency to the BMP

In the same way that Intent is an input to the phonetic function in §4.2, frequency information is yet another input. Frequency is represented as a function  $F$ , and the phonetic function is updated accordingly:  $A_{BP} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$ . In other words, phonetic implementation is a function that takes in the lexicon, the phonology, an intent variable and a frequency function. The frequency function we envision has the type  $F :: L \rightarrow \mathbb{R}$ . Since the lexicon  $L$  is a set, the frequency function maps each item in the lexicon to a number that corresponds to its frequency. Again, the inclusion of the input form of lexical items *vis-à-vis* the lexicon is what allows us to account for phonetic variation.

Furthermore, it is important that the same phonological form does not entail the same lexical item, since they are distinguished by syntactic and semantic information in the lexicon.

Another way to think about this is through the analogy of a computer's memory system. Each lexical item would be represented in memory as a unique bit string. The memory system does not care about the content of what it is storing; it just has different values stored at different bit addresses. The lexicon can be thought of in this same way. Under this type of architecture, the frequency information for a given lexical item is determined by a function rather than stored directly in the lexical entry. We see this as a way to encode the difference between knowledge *of* language and knowledge *about* language. The former refers to grammatical knowledge, while the latter refers to language use. Based on the studies discussed in the previous section, it is clear that both are necessary for the production process.

Before continuing further, we introduce a function  $\pi :: (UR \mid SR) \rightarrow PR$  that converts objects of the type  $UR$  or  $SR$  into a  $PR$ . Here, we assume this is a tuple of ordered cue parameter vectors. These may be articulatory or acoustic cues as long as they contain both spatial and temporal information. Given  $\pi$ , formula (10) discussed in the previous section for the implementation of the intent scaling would now be (13):

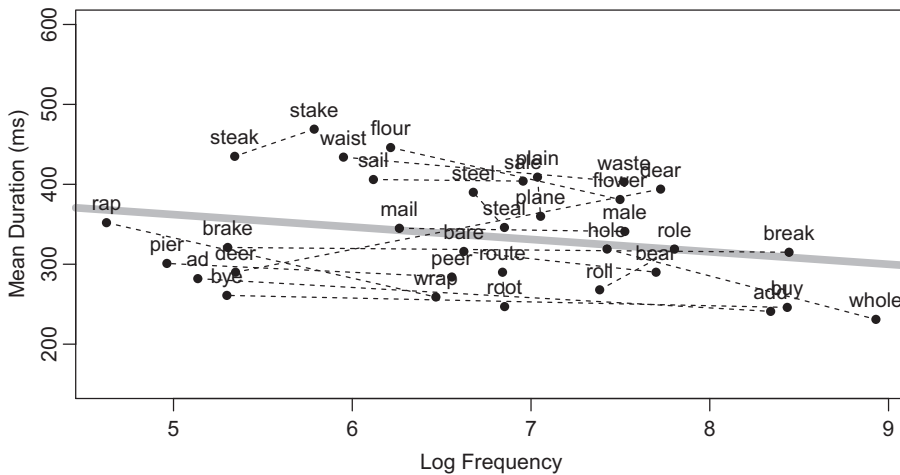
$$(13) \quad \tau = \pi(L) \times I^\alpha + \pi(P(L)) \times (1 - I^\alpha)$$

Recall that  $L$  contains  $UR$ s and  $P(L)$  returns  $SR$ s. So this is just the intent scaling over all cues for all phonemes of a given lexical item that is being produced. Here,  $\tau$  can be thought of as determining the overall target value with type  $\tau :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$ . It therefore provides a foundation that other factors can slightly alter. With that idea in mind, consider a duration scaling factor  $\delta :: \mathbb{R} \rightarrow [0, 1]^n$ . Specifically,  $\delta$  maps frequencies to the unit interval. These functions  $\pi$ ,  $\tau$  and  $\delta$  can be considered subprograms within the larger phonetic function  $A_{BP}$ .

In order to complete our description of this process, we also need to explain how the various input elements interact. In this simulation, we propose that the target value output by the  $\tau$  function is multiplied by the output of the  $\delta$  function to provide a frequency-scaled phonetic output. Following the assumption that the  $PR$  is a vector of parameters, the  $\delta$  function outputs a vector rather than a scalar. In this way, frequency effects can occur under the architecture of the BMP without needing to be encoded directly in the lexicon or the phonological grammar. Instead, they are just one more factor that influences production alongside the lexicon, the phonological grammar and pragmatic intent.

Our particular implementation is inspired by Pierrehumbert's (2002) simulation of leniting bias. She defines the production of a given token  $x$  as  $x = x_{target} + \varepsilon + \lambda$ , where  $x_{target}$  is the specific phonetic target that has been computed based on an exemplar model,  $\varepsilon$  is some random error, and  $\lambda$  is the leniting bias. This is motivated because leniting bias is closely related to duration (Priva & Gleason 2020), and duration is related to frequency. For our implementation, the equivalent of  $x_{target}$  is the output of  $\tau(L, P, I)$ , the equivalent of  $\lambda$  is the output of  $\delta(F(L))$ , and instead of adding the bias term to the target, our implementation multiplies them.

While the data we model only involves temporal cues, our implementation would equally apply to spectral cues as well. This raises the question of whether frequency information can also influence non-temporal cues. The answer appears to be yes. In a recent review of phonetic reduction, Clopper & Turnbull (2018) discuss how various factors such as frequency affect both spectral and temporal cues. The primary spectral cue that has been investigated in relation to frequency is the  $F1 \times F2$  vowel space, which has been shown to be more contracted for more frequent words (Munson & Solomon 2004). Crucially, Munson & Solomon (2004) found vowels in low-frequency words to be longer than vowels in high-frequency words, but found no statistically significant interaction between duration and vowel-space expansion. Therefore, a simulation that accounts for both spectral and temporal cues would necessarily have to tease apart the influence of duration from the influence of frequency. This, however, has no impact on the architecture of the BMP, since it already claims that both those types of information are available during the production process.



**Figure 6.** Average duration and log frequency for 17 homophonous pairs. These data come from the Switchboard corpus (Godfrey *et al.* 1992). Thin dashed lines connect all pairs. The thick grey line is the output of a linear model ( $\mathcal{I}$ ) of these points, showing a general negative correlation.

### 5.3. Homophone duration variation in English

In this section, we present a simulation that shows how the functions described in the previous section may be implemented using frequency data from CELEX (Baayen *et al.* 1996) and duration data from the Switchboard corpus (Godfrey *et al.* 1992). We gathered this data following the methodology presented by Gahl (2008: 479–480), including using the time-aligned orthographic transcript originally created by Deshmukh *et al.* (1998). Figure 6 shows the mean duration and log frequency of 17 homophonous pairs. Each point represents a word in the corpus and is connected to its homophone by a dashed line. While there is an overall negative correlation between duration and log frequency in the plotted pairs, it is not the case that every individual pair showed a negative relationship.<sup>12</sup>

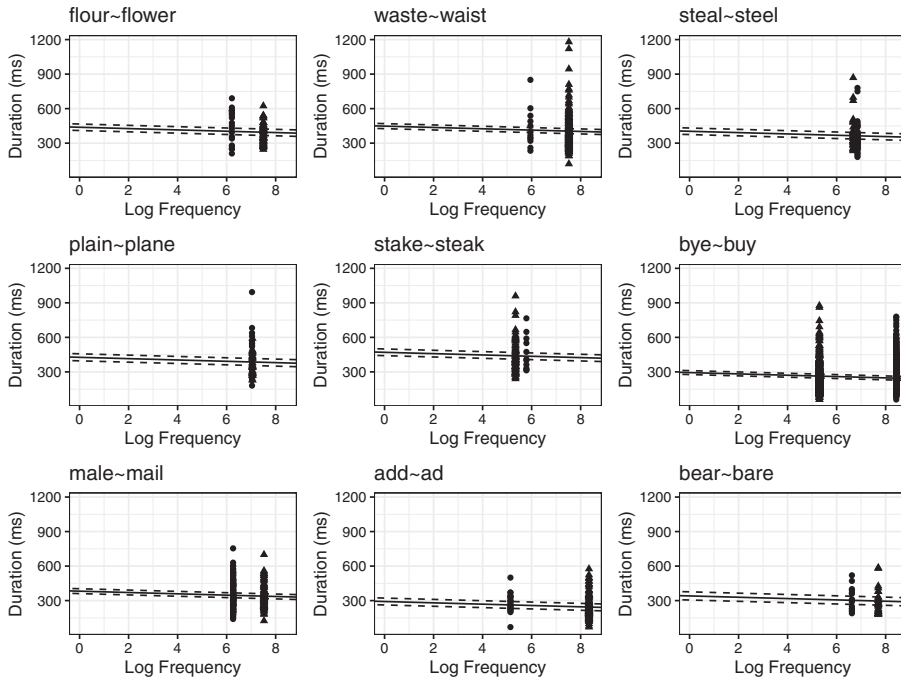
This simulation uses a linear model to predict the effect of frequency on duration. In the remainder of this section, it will be referred to as  $\mathcal{S}$  to reinforce the difference between the abstract model and this specific implementation.  $\mathcal{S}$ 's outcome variable ( $y$ ) is duration (in ms) and has two predictor variables: log frequency and phonological form. This results in a single slope based on log frequency and varying intercepts based on phonological form, and can be directly related to the functions for determining duration-influenced phonetic output. (14) shows the structure of  $\mathcal{S}$  in full.

$$(14) \quad y = \beta_0 + \beta_1 \times \text{LogFreq}(x) + \sum_{i=1}^{|L|} \beta_i \times [l_i = x] + \varepsilon$$

These parameters can be broken down to show how they relate to the functions above. Under the operating assumption that duration scales linearly with frequency, the underlying target value, which corresponds to the function  $\tau(L, P, I)$ , will be equal to the equation in (14) with  $\beta_1 \times \text{LogFreq}(x)$  removed. In other words, the intercept for each phonological form is the hypothesised target value.<sup>13</sup>

<sup>12</sup>Three of the most pronounced positive relationships all contain words where the same spelling results in different lexemes. For example, *deer* and *dear* have a large variation in frequency and a positive duration relationship. Following Gahl (2008), we collapsed words with the same spelling due to the difficulty of teasing apart meaning from orthography alone. Orthographic *dear* can stand for the noun or the adjective. A closer analysis may show that splitting these forms apart yields duration and frequency values that do follow the general trend. This is beyond the current scope of the article.

<sup>13</sup>A reviewer points out that since the  $\beta$  weight on frequency is a free parameter that is fitted to the data, there is nothing restricting the directionality of the effect. We agree this is a weakness of this simulation. Previous work has related the direction of the effect to exemplar storage (Gahl 2008) or motor practice (Bybee 2001). Another possibility is ordering the lexicon by frequency and implementing access as a linear search (cf. Yang 2016). In this case, more frequent words are shorter because they are accessed more quickly. This would also account for the direction of the effect. As discussed in §1, the BMP, as a computational



**Figure 7.** Frequency and duration information for individual tokens of 9 randomly selected homophonous pairs. Each plot represents a single pair. The solid black lines are the predicted linear relationship for that phonological form. Dashed lines indicate 95% confidence intervals.

To relate  $S$  to the duration-scaling function  $\delta(F(L))$  above, it is necessary to do some rearranging of terms. In its current form,  $S$  is similar to Pierrehumbert's (2002) approach. For the sake of exposition, replace  $\beta_0 + \sum_{i=1}^{|L|} \beta_i \times [l_i = x]$  with a constant  $k$  and remove the error term. The formula then becomes  $y = \beta_1 \times \text{LogFreq}(x) + k$ . Basic algebra derives an equivalent form:  $y = k \times (1 + \frac{\beta_1 \times \text{LogFreq}(x)}{k})$ . Since  $\beta_1$ , the slope coefficient, is negative and  $\text{LogFreq}(x)$  is guaranteed to be non-negative, the value of  $1 + \frac{\beta_1 \times \text{LogFreq}(x)}{k}$  is guaranteed to be less than 1. As long as  $\beta_1 \times \text{LogFreq}(x)$  is less than or equal to  $k$ , the value of  $1 + \frac{\beta_1 \times \text{LogFreq}(x)}{k}$  is also guaranteed to be greater than or equal to 0. Under these conditions, this works as a scaling factor in exactly the way necessary to implement the effect of frequency with the functions described above. The function  $\delta(F(L))$  above is therefore instantiated as (15).

$$(15) \quad 1 + \frac{\beta_1 \times \text{LogFreq}(x)}{\beta_0 + \sum_{i=1}^{|L|} \beta_i \times [l_i = x]}$$

Figure 7 shows the individual duration values for nine randomly selected homophonous pairs as well as the output of  $S$  for each phonological form.  $S$  has a significant effect for frequency ( $\beta = -5.761$ ,  $t = -3.561$ ,  $p < 0.001$ ). To illustrate how this works, consider the pair *bye~buy*.  $S$  predicts an intercept of 293.852 for this phonological form and therefore provides the equation  $\hat{y} = 293.852 - 5.761 \times \text{LogFreq}(x)$ . This can now be translated into the form  $PR_{dur} = \tau(L, P, I) \cdot \delta(F(L))$ . The term  $\tau(L, P, I)$  equals 293.852. For the form *bye*,  $\text{LogFreq}$  is equal to 5.30, making  $\delta(F(L))$  equal to  $(1 + \frac{-5.761 \cdot 5.3}{293.852}) = (1 + \frac{-30.5333}{293.852}) = (1 - 0.1039071) = 0.896$ . Using the same method,  $\delta(F(L))$  for *buy* is 0.835. These values therefore predict that the frequency-influenced duration value for *bye* should be  $293.852 \cdot 0.896 \approx 263$ . The mean duration for all tokens of *bye* in the data set is 261 ms. The frequency influenced duration value for *buy* is  $293.852 \cdot 0.835 \approx 245$ . The mean duration for all tokens of *buy* in the data set is 246 ms.

level description, has nothing to say about this issue and freely overgenerates. Other kinds of evidence or principles will be necessary to constrain it.



Success on an individual pair does not tell the entire story. To begin with, word frequency is not the only factor that affects duration. Second, the previous paragraph pairs the predicted value with the mean value for a given lexical item. Visual inspection of Figure 7 clearly shows that the data for each lexical item is quite spread. This suggests that the error term in  $\mathcal{S}$  can be directly thought of as the aspects of production other than frequency that influence duration for a given production. Therefore, specific results of  $\mathcal{S}$  presented here should be interpreted conservatively.

Rather than focussing on perfect prediction, the goal here was to show how the architecture of the BMP can be used to simulate this type of frequency and duration data. The assumptions being made in this simulation are: 1) the phonology maps discrete inputs to discrete outputs; 2) there are multiple inputs to the phonetic module: the target lexical item, the phonological map, the intent value and frequency information; 3) the lexical item, phonological map and intent are used to produce a PR; 4) this representation is further scaled based on frequency information for individual lexical items. Consequently, adopting an exemplar model or gradient phonology is not necessary to account for the types of duration effects that Gahl (2008) and others have documented.

## 6. Conclusion

This article introduced an abstract model of language production called the BMP, which is characterised in terms of typed functions. The crucial aspect of this model is that the phonetic production module is viewed as a higher-order function that takes the lexicon, phonology and other factors influencing production as its arguments. This view is contrasted with the standard modular feed-forward view, which describes the input to the phonetic production module as the output of phonology (Pierrehumbert 2002). Furthermore, we have demonstrated how this type of architecture can account for incomplete neutralisation, some cases of near-merger, and durational variation in homophones while maintaining discrete phonological knowledge.

The final type given to the phonetic production function is  $A_{BP} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$ . As discussed in §3.3, this is a curried function. What this means is that the lexicon, phonology, intent and frequency are all inputs to the function, and these arguments can be given one at a time. A function of arity  $n$  is said to be saturated if it has received  $n$  arguments. This perspective allows for the description of a chain of partially saturated production functions:

- (16) a.  $A_{BP} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$   
 b.  $A_{BP}^l :: P \rightarrow I \rightarrow F \rightarrow \{PR\}$   
 c.  $A_{BP}^{l,p} :: I \rightarrow F \rightarrow \{PR\}$   
 d.  $A_{BP}^{l,p,i} :: F \rightarrow \{PR\}$

These functions can be interpreted such that (16b) is the production function given a specific lexicon  $l$  in the set of all possible lexicons  $L$ ; (16c) is the production function given a specific lexicon and a specific phonology function  $p$  in the set of all possible phonology functions  $P$ ; and (16d) is the production function given a specific lexicon and phonology, as well as a specific intent value  $i$  in the set of all possible intent values  $I$ .

Consider another possible type,  $A'_{BP} :: (L, P) \rightarrow (I, F) \rightarrow \{PR\}$ . Here, the inputs are split into two tuples, one containing the lexicon and phonology and one containing the intent and frequency. This essentially can be viewed as the split between knowledge of language and knowledge about language. Since the act of production involves many factors beyond what has been discussed in this article, it is possible to switch  $(I, F)$  to a cover type  $E$  which stands in for all the information that goes into the production process other than the lexicon and phonology. With this in mind, it is possible to have a partially saturated function with type  $A'_{BP} :: E \rightarrow \{PR\}$ . Ignoring  $E$  completely here would result in a set of phonetic outputs influenced only by the lexicon and phonology.

Why does this matter? While it may appear that the phonetic module has been complicated by adding extra material to its input (the lexicon, intent, frequency), we argue instead that it has been simplified.

Typed functions allow for the larger production process to be broken down into its smaller pieces. What looks like a complicated system is instead the interaction of many different simple systems. In this way, type analysis is a new tool by which one can better understand the relationship between phonetics and phonology.

One consequence of this simplicity is that the BMP may appear too flexible, allowing all kinds of interactions that are not manifest in the phonetics–phonology interface. In general, models of the phonetics–phonology interface will have the same flexibility due to the level of analysis at which they are couched. For example, the feed-forward model itself is similarly ‘too flexible’. Nonetheless, this level of analysis still allows one to contrast the capacities and properties of different models. For example, as we have shown, the BMP alleviates problems inherent to the feed-forward model. Any particular theory of the interface will necessarily constrain the possibilities in some significant way. A reviewer asks what kind of criteria would be used to constrain the BMP. The answer is evidence from any scientific investigation can be brought to bear upon this question. For instance, we have reviewed in this article careful phonetic experimentation which has yielded evidence for the importance of extragrammatical factors on production. Additionally, other experimental work has shown the importance of maintaining categorical phonological knowledge (Du & Durvasula 2022; Mai *et al.* 2022). Considering van Rooij & Baggio’s (2020) characterisation of experimental and theoretical cycles in scientific research, our proposal can be thought of as a response to an experimental cycle dominated by the feed-forward model of the interface. The proposal in this article takes a step towards a new theoretical cycle, which can then lead to a new experimental cycle conducted within the perspective offered by the BMP.

Additionally, the BMP highlights the importance of certain kinds of information over others during the production process. While each factor plays a role in determining the phonetic output, the long-term memory representation of the pronunciation of a lexical item is arguably the most important factor, since the entire goal of the production process is to externalise it in some way. Phonology is also important, since it is largely viewed as an automatic process that systematically adjusts category level aspects of the pronunciation in a context-dependent way.<sup>14</sup> On the other hand, while pragmatic intent and lexical frequency systematically influence the phonetic output, they do so by scaling the targets that are determined by the lexicon and phonology.

This can also be related to a blueprint metaphor. Imagine there is a blueprint for building a picnic table. In one scenario a person uses this blueprint to build a table for an indoor area. In a second scenario, a different person uses the same blueprint to build a table to be used in an outdoor area. They both use the same materials and the same set of tools and end up with two tables that are practically identical. The person in scenario two then adds a clear coat of waterproofing since the table will be kept outside. To the naked eye there are still two identical tables, but closer inspection shows there is a fine-grained difference between the two. The blueprint is not explicit about how the table is used and therefore does not supply any further information beyond how to assemble the table. In spite of this, sometimes there are factors beyond its construction that affect its final form.

A reviewer asks how the BMP might handle gradient phonological phenomena that don’t arise external to the grammar. We reiterate that what we have shown in this article is that phenomena like incomplete neutralisation and systematic variation in homophone durations don’t necessarily require gradient phonological knowledge. What we have not shown (or argued for) is that phonological knowledge must necessarily be discrete. Our primary goal is not to assert that gradient phonological phenomena do not exist, but rather to highlight the fact that gradient measurements do not automatically imply gradient knowledge, since there may be alternate ways to account for this gradience (such as with the structure of the interface).

<sup>14</sup>We recognise that certain processes are optional and/or gradient, but would argue that phonological implementations of them still apply automatically. In other words, the optionality and gradience are determined by the automatic application of the phonology function.

The role of phonetics in the BMP is to take a set of materials (the lexicon) and a blueprint (the phonology) and construct the correct forms. Depending on the use of these forms, they are further altered by situational need (pragmatic context, frequency counts) to provide the final set of instructions to the motor system. In this sense, the BMP provides a phonologically based phonetics (cf. Hayes *et al.* 2004). The phonetic form is dependent on the phonological output, but there is plenty of room for systematic influence from other factors. In fact, the BMP in many ways is a formalised version of what Du & Durvasula (2022) call the ‘classic generative phonology’ view, which explicitly situates phonology as only one source of information in the production process. A clear description of this view comes from Mohanan (1986: 183, emphasis original):

Practitioners of phonology often distinguish between *internal* evidence, which consists of data from distribution and alternation, and *external* evidence, which consists of data from language production, language comprehension, language acquisition, psycholinguistic experiments of various kinds, sound patterning in versification, language games, etc. [...] The terms ‘internal’ and ‘external’ evidence indicate a bias under which most phonological research is being pursued, namely, the belief that the behaviour of speakers in making acceptability judgments is somehow a more direct reflection of their linguistic knowledge than their behaviour in producing language, understanding language, etc. This bias appears to be related to the fact that linguistic knowledge is only *one* of the inputs to language production, language comprehension, and other forms of language performance. What accounts for the facts of performance is a *conjunct* of a theory of linguistic knowledge (‘What is the nature of the representation of linguistic knowledge?’) and a theory of language performance (‘How is this knowledge put to use?’).

We believe the type analysis of the BMP provided in this article, along with simulations in our case studies, provides multiple entry points for further investigation of the BMP on its own terms or in comparison to other models of the interface. We began with some comparison with the research using dynamical systems because of its significant influence on our thinking. A reviewer points out that Jurafsky *et al.* (2002, Figure 3) and Shaw & Tang (2023) are other possible examples of research that could be instantiations of the BMP. The BMP also makes predictions in regard to the phonetic realisations of other kinds of phenomena including deletion, the realisation of absolutely neutralised segments, morphological boundary effects, and optionality.

In this article, we formalise the BMP using typed functions and show how the BMP architecture allows for the simulation of systematic phonetic gradience found in incomplete neutralisation, near-merger and homophone duration variation while maintaining a categorical phonological grammar. These simulations show that gradience within phonology, either in the representations or in the mappings, is not necessary to account for these types of data. This is not to say that phonology must be discrete and categorical, but rather that arguments against a discrete, categorical phonology based on incomplete neutralisation and similar phenomena are insufficient given the architecture of the BMP. As a result, the bound around what type of data the phonological grammar must account for has become tighter.

**Acknowledgements.** We thank Ellen Broselow, Karthik Durvasula, Marie Huffman, Bill Idsardi, the Stony Brook Mathematical Linguistics Reading Group and audiences at the Workshop on Theoretical Phonology 2020 and AMP 2021 for helpful discussions regarding this material. We also thank the *Phonology* reviewers and associate editor for their valuable comments which helped us improve the article significantly.

**Competing interests.** The authors declare no competing interests.

## References

- Albright, Adam & Bruce Hayes (2006). Modelling productivity with the gradual learning algorithm: the problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in grammar*. Oxford: Oxford University Press, 185–204.

- Archangeli, Diana (1984). *Underspecification in Yawelmani phonology and morphology*. PhD dissertation, MIT.
- Avery, Peter & William J. Idsardi (2001). Laryngeal dimensions, completion and enhancement. In T. Alan Hall (ed.) *Distinctive feature theory*. Berlin: de Gruyter, 41–70.
- Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers (1996). *The CELEX lexical database*. Philadelphia, PA: University of Pennsylvania. CD-ROM.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* **60**, 92–111.
- Bermúdez-Otero, Ricardo (2007). Diachronic phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press, 497–517.
- Bloomfield, Leonard (1933). *Language history: from Language*. New York: Holt, Rinehart and Winston.
- Braver, Aaron (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology* **36**, 1–36.
- Braver, Aaron & Shigeto Kawahara (2016). Incomplete neutralization in Japanese monomoraic lengthening. In Adam Albright & Michelle A. Fullwood (eds.) *Proceedings of the 2014 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America, 12 pp.
- Broselow, Ellen (1984). Default consonants in Amharic morphology. *MIT Working Papers in Linguistics* **7**, 15–31.
- Browman, Catherine P. & Louis Goldstein (1992). Articulatory phonology: an overview. *Phonetica* **49**, 155–180.
- Bybee, Joan (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Chandlee, Jane (2014). *Strictly local phonological processes*. PhD dissertation, University of Delaware.
- Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.
- Church, Alonzo (1932). A set of postulates for the foundation of logic. *Annals of Mathematics* **33**, 346–366.
- Church, Alonzo (1933). A set of postulates for the foundation of logic (second paper). *Annals of Mathematics* **34**, 839–864.
- Clopper, Cynthia G. & Rory Turnbull (2018). Exploring variation in phonetic reduction: linguistic, social, and cognitive factors. In Francesco Cangemi, Meghan Clayards, Oliver Niebuhr, Barbara Schuppler & Margaret Zellers (eds.) *Rethinking reduction: interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: Mouton de Gruyter, 25–72.
- Coats, Herbert S. & Alex P. Harshenin (1971). On the phonological properties of Russian. *The Slavic and East European Journal* **15**, 466–478.
- Coetzee, Andries W. & Joe Pater (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *NLLT* **26**, 289–337.
- Cohn, Abigail C. (1993). Nasalisation in English: phonology or phonetics. *Phonology* **10**, 43–81.
- Cohn, Abigail C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory* **16**, 1–31.
- Coleman, John & Janet Pierrehumbert (1997). Stochastic phonological grammars and acceptability. In John Coleman (ed.) *Computational phonology: third meeting of the ACL Special Interest Group in Computational Phonology*. Somerset, NJ: Association for Computational Linguistics, 49–56.
- Cooper, Richard P. & Olivia Guest (2014). Implementations are not specifications: specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research* **27**, 42–49.
- Cummins, Robert (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press.
- Curry, Haskell B. (1980). Some philosophical aspects of combinatory logic. In Jon Barwise, H. Jerome Keisler & Kenneth Kunen (eds.) *The Kleene symposium, number 101 in Studies in Logic and the Foundations of Mathematics*. Amsterdam: North-Holland Publishing Company, 85–101.
- Deshmukh, Neeraj, Aravind Ganapathiraju, Andi Gleeson, Jonathan Hamaker & Joseph Picone (1998). Resegmentation of SWITCHBOARD. In *Proceedings of the 5th International Conference on Spoken Language Processing*, paper no. 0685.
- Dinnsen, Daniel & Maria Garcia-Zamor (1971). The three degrees of vowel length in German. *Research on Language & Social Interaction* **4**, 111–126.
- Dinnsen, Daniel A. & Jan Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. *JPh* **12**, 49–60.
- Dmitrieva, Olga, Allard Jongman & Joan Sereno (2010). Phonological neutralization by native and non-native speakers: the case of Russian final devoicing. *JPh* **38**, 483–492.
- Du, Naiyan & Karthik Durvasula (2022). Phonetically incomplete neutralisation can be phonologically complete: evidence from Huai'an Mandarin. *Phonology* **39**, 559–595.
- Duhem, Pierre (1954). *The aim and structure of physical theory*. Princeton, NJ: Princeton University Press. Translated by Philip P. Wiener.
- Ernestus, Mirjam (2011). Gradience and categoricity in phonological theory. In Marc van Oostendorp, Colin J. Ewen, Keren Rice & Elizabeth V. Hume (eds.) *The Blackwell companion to phonology*, volume 4, chapter 89. Oxford: Wiley-Blackwell, 2115–2136.
- Ernestus, Mirjam & R. Harald Baayen (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Lg* **79**, 5–38.
- Feldman, Jerome A. & Dana H. Ballard (1982). Connectionist models and their properties. *Cognitive Science* **6**, 205–254.

- Feyerabend, Paul (1965). Reply to criticism: comments on Smart, Sellars and Putnam. In Robert S. Cohen & Marx W. Wartofsky (eds.) *Proceedings of the Boston Colloquium for the Philosophy of Science, 1962–1964*. Dordrecht: Springer, 223–261.
- Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* **18**, 7–44.
- Fodor, Jerry A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese* **28**, 97–115.
- Fodor, Jerry A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fosler-Lussier, Eric & Nelson Morgan (1998). Effects of speaking rate and word frequency on conversational pronunciations. In *Proceedings of Modeling Pronunciation Variation*, 35–40.
- Fougeron, Cécile & Donca Steriade (1997). Does deletion of French SCHWA lead to neutralization of lexical distinctions? In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, 943–946.
- Frrixione, Marcello (2001). Tractable competence. *Minds and Machines* **11**, 379–397.
- Gafos, Adamantios I. (2006). Dynamics in grammar: comment on Ladd and Ernestus & Baayen. *Laboratory Phonology* **8**, 51–80.
- Gafos, Adamantios I. & Stefan Benus (2006). Dynamics of phonological cognition. *Cognitive Science* **30**, 905–943.
- Gafos, Adamantios I., Simon Charlow, Jason A. Shaw & Philip Hoole (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *JPh* **44**, 152–166.
- Gahl, Susanne (2008). *Time and thyme* are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Lg* **84**, 474–496.
- Gerfen, Chip (2002). Andalusian codas. *Probus* **14**, 247–277.
- Godfrey, John J., Edward C. Holliman & Jane McDaniel (1992). SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*. New York: IEEE, 517–520.
- Goldrick, Matthew (2000). Turbid output representations and the unity of opacity. *NELS* **30**, 231–245.
- Gordon, Matthew (2004). Syllable weight. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*, chapter 9. Cambridge: Cambridge University Press, 277–312.
- Gorman, Kyle & Richard Sproat (2021). *Finite-state text processing*. San Rafael, CA: Morgan & Claypool.
- Gouskova, Maria & Nancy Hall (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In Steve Parker (ed.) *Phonological argumentation: essays on evidence and motivation*. London: Equinox, 203–225.
- Guest, Olivia & Andrea E. Martin (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior* **6**, 213–227.
- Gussmann, Edmund (2007). *The phonology of Polish*. Oxford: Oxford University Press.
- Hale, Mark & Charles Reiss (2000). “Substance abuse” and “dysfunctionalism”: current trends in phonology. *LI* **31**, 157–169.
- Hale, Mark & Charles Reiss (2008). *The phonological enterprise*. Oxford: Oxford University Press.
- Hall, Nancy (2013). Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *JPh* **41**, 133–143.
- Halle, Morris & Kenneth Stevens (1971). *A note on laryngeal features*. Quarterly Progress Report 101, Research Laboratory of Electronics, MIT.
- Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Heinz, Jeffrey (2018). The computational nature of phonological generalizations. In Larry M. Hyman & Frans Plank (eds.) *Phonological typology*. Berlin: de Gruyter Mouton, 126–195.
- Hinton, Geoffrey E. & James A. Anderson (eds.) (1989). *Parallel models of associative memory*. Updated edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Iverson, Gregory K. & Joseph C. Salmons (1995). Aspiration and laryngeal representation in Germanic. *Phonology* **12**, 369–396.
- Jassem, Wiktor & Lutosława Richter (1989). Neutralization of voicing in Polish obstruents. *JPh* **17**, 317–325.
- Jurafsky, Daniel, Alan Bell & Cynthia Girand (2002). The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*. Berlin: de Gruyter Mouton, 3–34.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Joan L. Bybee & Paul J. Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins, 229–254.
- Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Victoria A. Fromkin (ed.) *Phonetic linguistics: essays in honor of Peter Ladefoged*. Orlando: Academic Press, 115–132.
- Keating, Patricia A. (1988). Underspecification in phonetics. *Phonology* **5**, 275–292.
- Keating, Patricia A. (1990). Phonetic representations in a generative grammar. *JPh* **18**, 321–334.
- Kingston, John (2019). The interface between phonetics and phonology. In William F. Katz & Peter F. Assmann (eds.) *The Routledge handbook of phonetics*. New York: Routledge, 359–400.
- Kingston, John & Randy L. Diehl (1994). Phonetic knowledge. *Lg* **70**, 419–454.
- Kopkalli, Handan (1994). *A phonetic and phonological analysis of final devoicing in Turkish*. PhD dissertation, University of Michigan.
- Labov, William, Malcah Yaeger & Richard Steiner (1972). *A quantitative study of sound change in progress*, volume 1. Philadelphia, PA: U.S. Regional Survey.



- Lakatos, Imre (1970). Falsification and the methodology of scientific research programmes. In Imre Lakatos & Alan Musgrave (eds.) *Criticism and the growth of knowledge: proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge: Cambridge University Press, 91–196.
- Lambert, Dakotah Jay, Jonathan Rawski & Jeffrey Heinz (2021). Typology emerges from simplicity in representations and learning. *Journal of Language Modelling* 9, 151–194.
- Leslau, Wolf (1969). Frequency as determinant of linguistic changes in the Ethiopian languages. *Word* 25, 180–189.
- Lionnet, Florian (2017). A theory of subfeatural representations: the case of rounding harmony in Laal. *Phonology* 34, 523–564.
- Lohmann, Arne (2018a). *Cut* (n) and *cut* (v) are not homophones: lemma frequency affects the duration of noun–verb conversion pairs. *JL* 54, 753–777.
- Lohmann, Arne (2018b). *Time* and *thyme* are not homophones: a closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Lg* 94, e180–e190.
- Losiewicz, Beth (1995). Word frequency effects on the acoustic duration of morphemes. *JASA* 97, 32–43.
- Łukaszewicz, Beata (2021). The dynamical landscape: phonological acquisition and the phonology–phonetics link. *Phonology* 38, 81–121.
- Mai, Anna, Stephanie Riès, Sharona Ben-Haim, Jerry Shih & Timothy Gentner (2022). Phonological contrasts are maintained despite neutralization: an intracranial EEG study. In Peter Jurgec, Liisa Duncan, Emily Elfner, Yoonjung Kang, Alexei Kochetov, Brittney K. O'Neill, Avery Ozburn, Keren Rice, Nathan Sanders, Jessamyn Schertz, Nate Shaftoe & Lisa Sullivan (eds.) *Proceedings of the 2021 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America, 12 pp.
- Marr, David (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.
- McCarthy, John J. (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. London: Equinox.
- McCarthy, John J. & Alan S. Prince (1994). The emergence of the unmarked: optimality in prosodic morphology. *NELS* 24, 333–379.
- McCloskey, Michael (1991). Networks and theories: the place of connectionism in cognitive science. *Psychological science* 2, 387–395.
- Mohanan, Karuvannur Puthanveetil (1986). *The theory of Lexical Phonology*. Dordrecht: D. Reidel.
- Munson, Benjamin & Nancy Pearl Solomon (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048–1058.
- Myers, Scott (2000). Boundary disputes: the distinction between phonetic and phonological sound patterns. In Noel Burton-Roberts, Philip Carr & Gerard Docherty (eds.) *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press, 245–272.
- Newell, Allen & Herbert A. Simon (1958). Elements of a theory of human problem solving. *Psychological Review* 65, 151.
- Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth (2018). Using meta-analysis for evidence synthesis: the case of incomplete neutralization in German. *JPh* 70, 39–55.
- Ohala, John J. (1983). The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.) *The production of speech*. New York: Springer, 189–216.
- Ohala, John J. (1990). There is no interface between phonology and phonetics: a personal view. *JPh* 18, 153–172.
- Ohala, John J. (1992). The costs and benefits of phonological analysis. In Pamela A. Downing, Susan D. Lima & Michael Noonan (eds.) *The linguistics of literacy*. Amsterdam: Benjamins, 211–238.
- van Oostendorp, Marc (2008). Incomplete devoicing in formal phonology. *Lingua* 118, 1362–1374.
- Pierce, Benjamin C. (2002). *Types and programming languages*. Cambridge, MA: MIT Press.
- Pierrehumbert, Janet B. (1990). Phonological and phonetic representation. *JPh* 18, 375–394.
- Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology* 7. Berlin: de Gruyter Mouton, 101–140.
- Popper, Karl (1959). *The logic of scientific discovery*. London: Routledge. Translated by the author from the original German.
- Port, Robert & Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *JPh* 17, 257–282.
- Port, Robert, Fares Mitleb & Michael O'Dell (1981). Neutralization of obstruent voicing in German is incomplete. *JASA* 70, S13.
- Port, Robert F. & Adam P. Leary (2005). Against formal phonology. *Lg* 81, 927–964.
- Port, Robert F. & Michael L. O'Dell (1985). Neutralization of syllable-final voicing in German. *JPh* 13, 455–471.
- Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Technical Report 2, Rutgers Center for Cognitive Science.
- Priva, Uriel Cohen & Emily Gleason (2020). The causal structure of lenition: a case for the causal precedence of durational shortening. *Lg* 96, 413–448.
- Pulleyblank, Douglas (1988). Underspecification, the feature hierarchy and Tiv vowels. *Phonology* 5, 299–326.
- Putnam, Hilary (1967). Psychological predicates. In William H. Capitan & Daniel Davy Merrill (eds.) *Art, mind, and religion*. Pittsburgh, PA: University of Pittsburgh Press, 37–48.
- Quine, Willard V. O. (1951). Two dogmas of empiricism. *Philosophical Review* 60, 20–43.
- Ramer, Alexis Manaster (1996). A letter from an incompletely neutral phonologist. *JPh* 4, 477–489.
- Reiss, Charles (2018). Substance free phonology. In S.J. Hannahs & Anna Bosch (eds.) *The Routledge handbook of phonological theory*. New York: Routledge, 425–452.



- Reiss, Charles & Veno Volenec (2022). Conquer primal fear: phonological features are innate and substance-free. *Canadian Journal of Linguistics* **67**, 581–610.
- Revithiadou, Anthi (2008). Colored turbid accents and containment: a case study from lexical stress. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?* Berlin: de Gruyter Mouton, 149–174.
- Roark, Brian & Richard Sproat (2007). *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.
- van Rooij, Iris (2008). The tractable cognition thesis. *Cognitive Science* **32**, 939–984.
- van Rooij, Iris & Giosuè Baggio (2020). Theory before the test: how to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science* **16**, 682–697.
- Roon, Kevin D. & Adamantios I. Gafos (2016). Perceiving while producing: modeling the dynamics of phonological planning. *Journal of Memory and Language* **89**, 222–243.
- van Rooy, Bertus, Daan Wissing & Dwayne D. Paschall (2003). Demystifying incomplete neutralisation during final devoicing. *Southern African Linguistics and Applied Language Studies* **21**, 49–66.
- Rumelhart, David E., James L. McClelland & the PDP Research Group (1988). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations*. Cambridge, MA: MIT Press.
- Saltzman, Elliot L. & Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* **1**, 333–382.
- Shaw, Jason Anthony & Kevin Tang (2023). A dynamic neural field model of leaky prosody: proof of concept. In Noah Elkins, Bruce Hayes, Jinyoung Jo & Jian-Leat Siah (eds.) *Proceedings of the 2022 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America, 12 pp.
- Slowiaczek, Louisa M. & Daniel A. Dinnsen (1985). On the neutralizing status of Polish word-final devoicing. *JPh* **13**, 325–341.
- Smolensky, Paul & Matthew Goldrick (2016). Gradient symbolic representations in grammar: the case of French liaison. Ms., Johns Hopkins University and Northwestern University. ROA #1286.
- Solé, Maria-Josep (1992). Phonetic and phonological processes: the case of nasalization. *Language and Speech* **35**, 29–43.
- Solé, Maria-Josep (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech* **38**, 1–23.
- Solé, Maria-Josep (2007). Controlled and mechanical properties in speech: a review of the literature. In Maria-Josep Solé, Patrice Speeter Beddor & Manjari Ohala (eds.) *Experimental approaches to phonology*. Oxford: Oxford University Press, 302–321.
- Strycharczuk, Patrycja (2019). Phonetic detail and phonetic gradience in morphological processes. In Mark Aronoff (ed.) *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.
- Tuller, Betty, Pamela Case, Mingzhou Ding & J. A. Scott Kelso (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* **20**, 3–16.
- Volenec, Veno & Charles Reiss (2017). Cognitive phonetics: the transduction of distinctive features at the phonology–phonetics interface. *Biolinguistics* **11**, 251–294.
- Walsh, Thomas & Frank Parker (1983). The duration of morphemic and nonmorphemic /s/ in English. *JPh* **11**, 201–206.
- Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *JPh* **32**, 251–276.
- Westbury, John R. & Patricia A. Keating (1986). On the naturalness of stop consonant voicing. *JL* **22**, 145–166.
- Wheeler, Max W. (2005). *The phonology of Catalan*. Oxford: Oxford University Press.
- Winter, Bodo & Timo Röttger (2011). The nature of incomplete neutralization in German: implications for laboratory phonology. *Grazer Linguistische Studien* **76**, 55–74.
- Yang, Charles (2016). *The price of productivity*. Cambridge, MA: MIT Press.
- Yu, Alan C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology* **24**, 187–214.
- Yu, Alan C. L. (2011). Mergers and neutralization. In Marc van Oostendorp, Colin J. Ewen, Keren Rice & Elizabeth V. Hume (eds.) *The Blackwell companion to phonology*, volume 3, chapter 80. Oxford: Wiley-Blackwell, 1892–1918.
- Zhang, Jie (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.) *Phonetically based phonology*. Cambridge: Cambridge University Press, 157–190.
- Zsiga, Elizabeth C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *JPh* **28**, 69–102.