# The blueprint model of production

Scott Nelson and Jeffrey Heinz*

**Abstract**

This paper introduces the blueprint model of production, which characterizes 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 phonetics module which takes the output of a phonological module as its input. The central idea of the blueprint model of production 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 neutralization, 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.

**Key Words:** phonetics-phonology interface, language production, computational phonology, typed functions, incomplete neutralization, homophone variation

# 1 Introduction

The division of labor 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 is blind to everything but the phonological output.

Certain types of phonetic data such as incomplete neutralization (Port et al., 1981; Port and O'Dell, 1985) and variation in durational properties of homophones (Gahl, 2008) are two 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,

*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 paper significantly.

1992; Pierrehumbert, 2002; Port and Leary, 2005). This often results in researchers reconceptualizing phonology by making it continuous or even eliminating the distinction between the phonetic and phonological modules altogether (Hayes et al., 2004; Browman and Goldstein, 1992; Zsiga, 2000, 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) that reconceptualizes how the modules formally interact. The formal characterization of this reconceptualization is described using typed functions (Pierce, 2002), which are related to the lambda calculus (Church, 1932, 1933). Under this view, the phonetics and phonology modules are functions.[1] 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 $g$: $f(g(x))$, but rather as the output of a higher order function that takes the phonology function alongside the lexical item as inputs: $f(g(), x)$. Furthermore, the phonetics module may take additional inputs $y$, such as speaker's intent to maintain an underlying contrast, as evidence warrants: $f(g(), x, y)$.

Viewing the phonetics module this way allows for information about both the underlying lexical form, as well as 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 neutralization 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 and underlying representations (Goldrick, 2000; Gafos and Benus, 2006; Van Oostendorp, 2008; Braver, 2019). As section 3 explains, the relationship of these proposals to each other and others is made clearer by the typed-function analysis 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 mistake a simulation with the higher level model, but as McCloskey (1991, p. 390) warns, "any simulation includes theory-irrelevant as well as theory-relevant details; hence, the details of a

[1]For the remainder of this paper the terms 'function' and 'module' are used interchangeably.

simulation cannot be identified straightforwardly with the details of the corresponding theory." Cooper and 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.

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 characterization 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 and Baggio, 2020), not specific implementations. This type of abstract analysis runs into the problem of *multiple realizability* (Putnam, 1967; Fodor, 1974; Guest and 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[2] 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

[2] `https://en.wikipedia.org/wiki/Sorting_algorithm`

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 paper is laid out as follows: §2 gives an overview of previous accounts of the relationship between phonetics and phonology. §3 provides the formalization 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 neutralization and variation in homophone durations. §4 focuses on final devoicing in German (Port and Crawford, 1989), tonal near merger in Cantonese (Yu, 2007), and epenthesis in Lebanese Arabic (Gouskova and Hall, 2009; Hall, 2013). This section further formalizes the relationship between incomplete neutralization 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 paper 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 and Halle, 1968), *SPE* is a natural starting point for the current discussion. In *SPE* it is assumed that the phonology contains rules that map binary features to a scalar value so that the surface representation (SR) of a lexical item is a temporally organized 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 or not both types of rules interact. Additionally, *SPE* assumes that there is a phonetics 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 and 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 and Leary's (2005) paper titled *Against Formal Phonology*. They argue that a discrete formal symbolic system is unable to account for the variability in phonetic realization of identical symbols as well as certain temporal based contrasts in behavioral data. Since these formal systems cannot simulate the natural language data on their own, Port and Leary (2005) argue against having a formal phonological grammar at all. Ohala (1990) takes a softer approach. He recognizes 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" (p. 156).

Two formal proposals that dissolve the distinction between phonetics and phonology are Flemming's (2001) unified model of phonetics and phonology and Browman and Goldstein's theory of Articulatory Phonology (1992, *et seq.*). Flemming (2001) develops an Optimality Theoretic (OT; Prince and 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 (Hz) and duration (ms).

Articulatory Phonology (AP; Browman and Goldstein, 1992) 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 and 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 and Benus, 2006; Gafos et al., 2014; Roon and Gafos, 2016; Łukaszewicz, 2021, among others).

If we reject the previous discussed accounts and instead favor 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 and 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 generalizations (Albright and Hayes, 2006; Ernestus, 2011). A second way regards grammatical acceptability judgments (Coleman and Pierrehumbert, 1997; Coetzee and Pater, 2008). A third way, a focus of this paper, is in relation to representations (Smolensky and Goldrick, 2016; Lionnet, 2017).

Beyond deciding which type of phonological gradience is applicable to the phonetics-phonology interface, Pierrehumbert (1990, p. 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 are either 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 $*\textsc{Dur}(\tau_i)$ 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 $*\textsc{Dur}(\tau_i) \gg *\textsc{Dur}(\tau_j)$. 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 showed 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 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 and 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 and 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 and Reiss (pp. 154–156; 2008) 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 is difficult to not 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 papers have clarified that it is only the phonological computations that are devoid of any substantive influence, but the phonological representations still have phonetic correlates (Volenec and Reiss, 2017; Reiss and Volenec, 2020). Volenec and 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 underlying and surface representations must both be binary phonological feature bundles. It is only through a separate *transduction* that any type of phonetic representation can be established. They posit a transducer which they refer to as "Cognitive Phonetics" which translates the output of phonology (an SR) into a phonetic representations (PR). The PR is "is a complex array of neural commands that activate muscles involved in speech production" (p. 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 and Benus (2006) who write "...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" (p. 924). 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 characterization of the phonetics-phonology interface based on typed functions (Pierce, 2002; Church, 1932, 1933). Our general characterization falls under the Marrian *computational level* category (Marr, 1982) as we follow van Rooij and Baggio's (2020) proposal for adopting a "top-down approach" in modeling 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 and Baggio, 2020, p. 684). Our more specific characterization 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 conceptualizing the phonetics production module as a higher-order function that takes the phonology module as an argument. This does away with "the metaphor of precedence" at the interface while maintaining a distinction between phonology and phonetics (Gafos and 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 reconceptualization 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 characterization 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.[3] Second, the whole phonological module is a factor in production, not just the representations it outputs. Like Gafos and 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 and 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 reconceptualizes 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 situate 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 neutralization.

## 3.1 Characterizing 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 specialized stations until it reaches the end point as a phonetic object that can be pronounced. Since assembly lines are successive in nature, each station is essentially blind to the history of the objects it receives. To make this metaphor more concrete, we can imagine that the Lexicon places Underlying Representations (URs) on a conveyor belt which takes them to the Phonology station to be worked on. At the Phonology station, URs are transformed into Surface Representations (SRs) and SRs are placed back on the conveyor belt to be taken down the line to the Phonetics station.

[3] The simultaneous, or parallel, view presented here may evoke connectionist models of cognition (Hinton and Anderson, 1981; Feldman and Ballard, 1982; Rumelhart et al., 1988). Our use of simultaneity varies from the connectionist view since 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 the functions we propose can be instantiated in any number of ways, including ones which follow connectionist/neural principles and ones that do not.

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 the Phonology was not a station that a lexical item had to pass through during the production process, but rather a target design that the phonetics 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 and Benus (2006)). This characterization 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 underlying representation. It also allows the Phonetics Station simultaneous access to both the underlying representation 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 evalaution is visible to the Phonetics Station.[4] 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 and Sproat, 2007; Heinz, 2018). In the modular feed-forward model, the phonology module is a function that maps a $UR$ to an $SR$ and the phonetics module is a function that maps an $SR$ to a $PR$ (Phonetic Representation). The BMP continues to view the phonology module as a function that maps a $UR$ to an $SR$ but views the phonetics module as a higher order function that *takes the phonology module function as an input*. In addition, to generalize over all lexical items, we consider the entire lexicon to be an input to the phonetics module instead of a single $UR$.[5] The phonetics module is therefore a function with at least two inputs: the lexicon and the phonology module; and one output: a set of phonetic representations $\{PR\}$. The two contrasting models are shown in Figure 1 below. The next section provides a formal definition of the BMP.

[4]The only way around this would be to encode such information into the SR itself. For example, in 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 the 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).

[5]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 and Sproat, 2007; Gorman and Sproat, 2021).

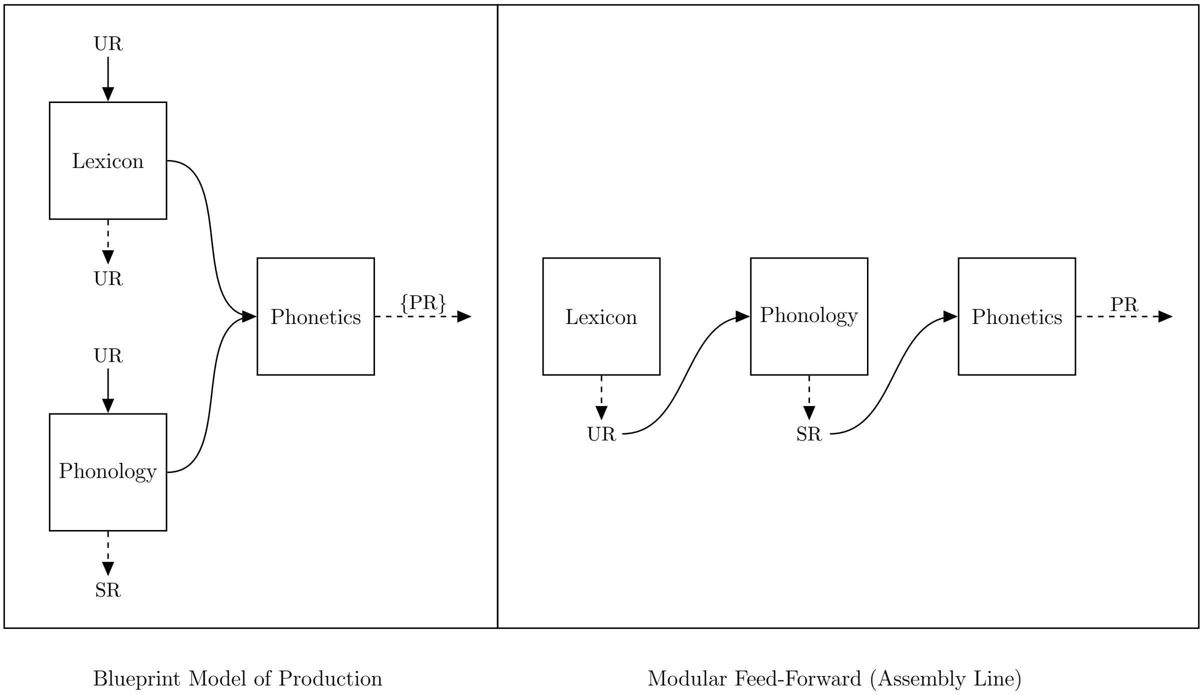


Figure 1: Visual comparison of the architecture for modular feed-forward models and the BMP. Each box represents a function/module. Solid lines represent the inputs to each function while dashed lines represent the outputs of each function.

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

While giving the phonetics 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 representations output by the phonological module *includes* the input to the phonological module (Prince and Smolensky, 1993; Goldrick, 2000; Van Oostendorp, 2008; Revithiadou, 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.

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 phonology 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 naumbers to a single real number. Later in section 3.3 we 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 $[x_1, \ldots, x_n]$, $map(f)(\vec{x})$ applies function $f$ to every individual element of $x \in \vec{x}$ and returns the array $[f(x_1), \ldots, f(x_n)]$. 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 summarize, 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 $function\text{-}application :: X \rightarrow (X \rightarrow Y) \rightarrow Y$. 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 and Phonetics (**A**rticulation or **A**coustics); $UR$, $SR$, and $PR$ to represent **U**nderlying **R**epresentations, **S**urface **R**epresentations, and **P**honetic **R**epresentations. The proposed types are listed in the table in Table 1.[6]

[6]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.

| Name | Meaning | Type |
|---|---|---|
| $L$ | Lexicon | $UR \to UR$ |
| $P$ | Phonology | $UR \to SR$ |
| $A_{\text{MFF}}$ | Phonetics$_{\text{MFF}}$ | $SR \to PR$ |
| $A_{\text{BP}}$ | Phonetics$_{\text{BP}}$ | $L \to P \to \{PR\}$ |
| $UR$ | Underlying Representation | $UR$ |
| $SR$ | Surface Representation | $SR$ |
| $PR$ | Phonetic Representation | $PR$ |

Table 1: Types

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

(1) $A_{\text{MFF}} :: SR \to PR$

This idealizes the phonetics module as a map from surface representations to phonetic representations. Given a UR, the phonology P, and the definition of function application from above, one can decompose $SR$ into $UR \to (UR \to SR)$.

(2) $A :: UR \to (UR \to SR) \to PR$

Next, $(UR \to SR)$ is just another way of representing the phonology module.

(3) $A :: UR \to P \to PR$

To complete this reconceptualization we change $UR$ to $L$ in order to generalize over the entire lexicon. By doing so, the output is now a set of phonetic representations rather than a single specific representation. This gives us a new type for the phonetics function.

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

The phonetics module is therefore a higher-order function with two arguments: the lexicon and the entire phonology 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 phonetic representation. The main difference is phonology is no longer an intermediary between the lexical form and the phonetics module. Instead, the phonology and the lexical form are both input to the phonetics module.

If it is not clear yet as to why this is being called the BMP, 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 which is 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 phonetics module as to how a given lexical item should be pronounced. In other words, in the same way one would query a blueprint, the phonetics 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 neutralization 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 extra linguistic factors relating to contrastive intent (Port and Crawford, 1989; Ernestus and Baayen, 2003; Gafos and 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 included its inputs (Prince and Smolensky, 1993; Goldrick, 2000; Van Oostendorp, 2008; Revithiadou, 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 and Smolensky, 1993). Under the feed-forward model, the principle of containment ensures that the phonetics 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 phonetics module is able to, for example, distinguish between *faithful*, word-final, voiceless obstruents and *derived* ones (Van Oostendorp, 2008).

Note 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 surface representation presented alongside the underlying representation 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 surface representation paired with an underlying representation. 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: $add :: \mathbb{R} \to \mathbb{R} \to \mathbb{R}$. This can be thought of as the uncurried version of $add' :: (\mathbb{R}, \mathbb{R}) \to \mathbb{R}$. 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* (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 $curry :: ((A, B) \to C) \to (A \to B \to C)$. 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 $P' :: UR \to (UR, SR)$. Under the feed-forward model, this output is given to the phonetics module to produce the articulatory representation. Consequently, the phonetics module would have type $A' :: (UR, SR) \to PR$. This is essentially the *uncurried* version of the BMP. Currying $A'$ yields a phonetics function of the form in (5).

(5) $curry(A') :: UR \to SR \to PR$

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

(6) $curry(A') :: P \to PR$

Combining (6) and (3) reveals that the BMP can be characterized as shown below.

(7) $A :: UR \to P \to PR$

Generalizing over the lexicon again, we get the same type for the BMP.

(8) $A_{\text{BPM}} :: L \to P \to \{PR\}$

This shows precisely the relation between containment theories of phonology under the feed-forward model and *any* theory of phonology computing functions $UR \to SR$ with the BMP. It also highlights the essential difference between the BMP and the modular feed-forward model: the latter serializes 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 neutralization (Port et al., 1981; Port and 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 auxillary 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 neutralization 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 Neutralization

This section first provides background on incomplete neutralization. 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 and Crawford, 1989), tonal merger in Cantonese (Yu, 2007), and vowel epenthesis in Lebanese Arabic (Gouskova and Hall, 2009; Hall, 2013).

## 4.1 Background

Final devoicing is probably the most well studied example of a phonological neutralization 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.[7] 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 and Harshenin, 1971), and Turkish (Kopkalli, 1994). The data in (9) provide an example from German (Dinnsen and Garcia-Zamor, 1971).

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

(9) a. /bad+en/ → [baden] 'to bathe'
b. /bad/ → [bat] 'bath'
c. bat+en/ → [baten] 'asked'
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 and 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 neutralization".

Final devoicing has been extensively studied, and found to be incomplete in many other languages such as Catalan (Dinnsen and Charles-Luce, 1984), Dutch (Warner et al., 2004), Polish (Slowiaczek and Dinnsen, 1985), Russian (Dmitrieva et al., 2010), and Afrikaans (Van Rooy et al., 2003). Many other processes, such as coda aspiration in Andalusian Spanish (Gerfen, 2002), French schwa deletion (Fougeron and Steriade, 1997), and Japanese monomoraic lengthening (Braver and 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 and Crawford (1989) find that listeners appear to have control over the level of incompleteness of the neutralization based on communicative context and how salient a contrast is made. In their experiment, they used five different contexts (based on 4 sentence conditions) to evaluate how the level of neutralization changed depending on speakers' awareness of the task. Condition 1A/B were 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. Condition 4 was the words in isolation.

They found incomplete neutralization in every condition when analyzing aggregated speaker data. No difference in the amount of incomplete neutralization was detected between conditions 1A and 1B in contrast to previous experiments (Jassem and Richter, 1989). In all other cases reported in Port and 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/B 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 amount 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 neutralized a segment is depending on the contrastive condition. The pragmatic conditions therefore influence a speakers intent on maintaining an underlying contrast. In their nonlinear dynamic approach to production, Gafos and Benus (2006) include a variable called *intent* to account for this fact. For the remainder of this paper, we will also use the term *intent* as a coverall term indicating pragmatic condition/desire to maintain an underlying contrast.

## 4.2 How the BMP includes Intent

Section 3.2 provided a formal characterization of the production process. The focus in that section was to show how the BMP conceptualizes the phonetics 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 neutralization. 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 be: $A_{\text{BP}} :: L \to P \to I \to \{PR\}$.

In other words, the inputs to the phonetics 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 phonetics module can account for incomplete neutralization 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 neutralization data while maintaining a discrete phonology. The one outlined below is used in the simulations in the remainder of the paper.

Recall that the acoustic cues in incompletely neutralized 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 neutralization. Additionally, Port and 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 being incomplete and why it is only a subset of cues lies beyond the scope of this paper. 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 a way such 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] sub-section 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] sub-section can be full of different realizations. 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.[8]

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 in which the scaling procedure may be implemented. Figure 2 provides a visual conceptualization of the cue space for the words in (9). Arrows point to possible realizations. Notice that it is only the alternating case where multiple options exist for a given form.

[8]We are assuming here that the phonetics module is able to map a [+voice] sound at the end of a word onto some phonetic representation. Since the translation is feature based this should not be a problem. The reason 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. Anecdotally, speakers of languages with final devoicing can produce a word final obstruent as voiced if absolutely forced to do so.

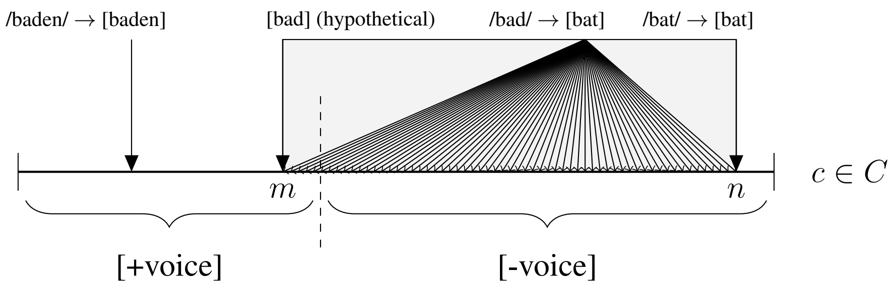


Figure 2: Hypothetical Cue space

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 $A_{\text{BP}}$ 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 formalize 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 $100\%$ maintain the underlying contrast. 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 neutralization 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) $c = c_{UR} \times I^{\alpha} + c_{SR} \times (1 - I^{\alpha})$

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 neutralization are subtle, and also that, under extreme circumstances, speakers can produce something very UR-like (see fn. 8).

## 4.3 Comparison to dynamical system approach

As discussed above, the approach of Gafos and 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, extra-grammatical 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 and 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 and Benus (2006).

Consider an instantiation of the BMP to cases where the type realizations for the underlying representations, the surface representations, 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 $\dot{x} = f(x) = -dV(x)/dx$ that describes a time-invariant first-order dynamical system in control of a cue, and $f(x)$ 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 phonetics module in the BMP could then combine them in the way as described in equation (10) above to get the final $PR$ form. Ultimately, Gafos and Benus (2006) take a different approach to their dynamics. They use a tilted anharmonic oscillator to formalize 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 $\theta$ parameter to express contrastive intent within their faithfullness 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 emphasize clear parallels between the BMP and the specific approach of Gafos and Benus (2006). Where the BMP associates a cue value with a UR, they 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 paper is that language production involves the interaction of lexical, phonological, and extra-grammatical factors which the modular feedforward model fails to capture. Since this idea is able to 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 behavior 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 neutralization 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 neutralization 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 behavior 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 and Crawford's (1989) results from German that show that the level of incompleteness can vary based on pragmatic factors. This sections 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 and Crawford (1989) found to be incomplete, and closure duration which they found to be complete. The exact cues found to incompletely neutralize has varied across studies. For example, there have been conflicing results about whether or not preceding vowel duration is an incomplete cue in German final devoicing. 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 neutralization. Port and Crawford (1989) on the other hand reported preceding vowel duration as being complete in their findings.

Our assumption, based on the results of Port and Crawford (1989), is that the level of incomplete neutralization 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 between Port and 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 neutralize incompletely or not, but rather on the working assumption that cues can vary in this way at all.

| 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 | |

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

Mean values for both closure duration and burst duration for each neutralized final stop pair and each condition are shown in Table 2. These data are taken directly from Port and Crawford (1989, Table 1; p. 265).[9] The ratio columns were added by dividing the final /d/ values from the final /t/ values within 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 within 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 significantly vary between derived and faithful surface /t/ segments in the pooled data, but closure duration (represented as $\times$) was not, the $\alpha$ parameter was set to 1 for burst duration and 20 for closure duration.

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.[10]

From the figures, it is possible to compare both within plots and between plots, resulting in four comparisons. Based on the Port and Crawford (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. Within the left plot, the cue values are shown to vary between burst duration ($+$) and closure duration ($\times$), as expected. Cue values close to 1 indicate that the final /d/ that has been neutralized 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

[9]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.

[10]In their discriminant analysis, Port and Crawford (1989) found that condition 2 was more easily recognized as underlying /d/ than conditions 1. This goes against the acoustic data presented in the paper 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.

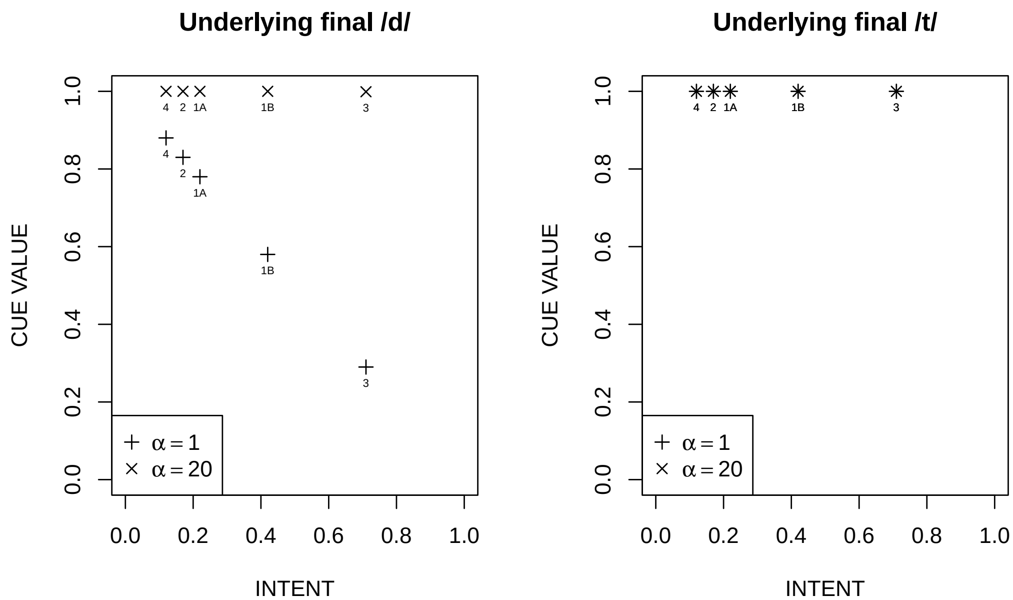


Figure 3: Simulated cue values for Port and Crawford (1989) results. Left plot shows values for /d/-final words and right plot shows values for /t/-final words. Symbols $+$ and $\times$ represent burst and closure duration, respectively.

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 behaviors observed by Port and Crawford (1989), where burst duration was incompletely neutralized and varied according to pragmatic context, while closure duration did not. A reviewer points out that closer inspection of individual behavior in Port and 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 behavior 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 behaviors, we don't believe that would provide any further insight into what we find important: the capacity of the individual 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 behavior the simulation exhibits.

The overall structure of the BMP allows for lexical influence on phonetic form. It also accounts for incomplete neutralization while maintaining a singular phonological devoicing rule, contra Port and 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 (p. 280)". This interpretation follows from conceptualizing the phonetics-phonology interface in terms of the modular feed-forward model, but it does not follow from conceptualizing 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 recognizing 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 and Crawford (1989) show that pragmatic context is a necessary ingredient, which is formalized 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 role 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 neutralization with near merger has been well documented (Ramer, 1996; Winter and Röettger, 2011; Yu, 2011; Braver, 2019). While the term incomplete neutralization emerged from the phonetics and phonological literature, the term near-merger originated within the sociolinguistic literature. 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. 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 neutralization 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 neutralization and near-merger is whether or not 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 neutralization is often related to the active production process.

Alternations also help distinguish the two. In the *source* vs. *sauce* example, there is no alternation driving the neutralization, but incomplete neutralization 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 neutralization using the BMP.

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

(11) a. sou33 'to sweep' → sou35 'a broom'
b. pɔŋ22 'to weigh' → pɔŋ35 'a scale'
c. tsʰɵɥ11 'to hammer' → tsʰɵɥ35 'a hammer'

The derived mid-rising tones of these *pinjam* words were compared with lexical mid-rising tones in lexical near-minimal pairs. The f0 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 a non-high level tone. A closer look shows that the phonological analysis involves an underlying floating high tone: pɔng22(55) → pɔng35 '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 f0 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 /tɐk˺/ both surface with a mid-rising tone in contracted syllables:

(12) a. paŋ22 tsɔ35 → pɔ35 'to weigh (PERF)'
b. pɔŋ22 tɐk˺55 → pɔ35 'to weigh (POTENTIAL)

What makes it interesting is that /tsɔ/ has an underlying mid-rising tone while /tɐk˺/ has an underlying high tone. The f0 value at all of the 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 underlying 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 f0 values raised, given a non-zero intent value. Despite the exemplar interpretation, Yu (2007, p. 207) recognizes this fact and writes, "Thus, the extra-high f0 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 "inbetween-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]). Squares 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 is therefore the black square.

In the simulation, the faithful mappings are unaffected by the $\alpha$ and intent values, and the values for the alternation mapping is 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 inbetween-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 "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).

Yu (2007) also found that the mid-rising tone derived from an underlying high tone in the contracted syllables had higher f0 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 neutralization 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

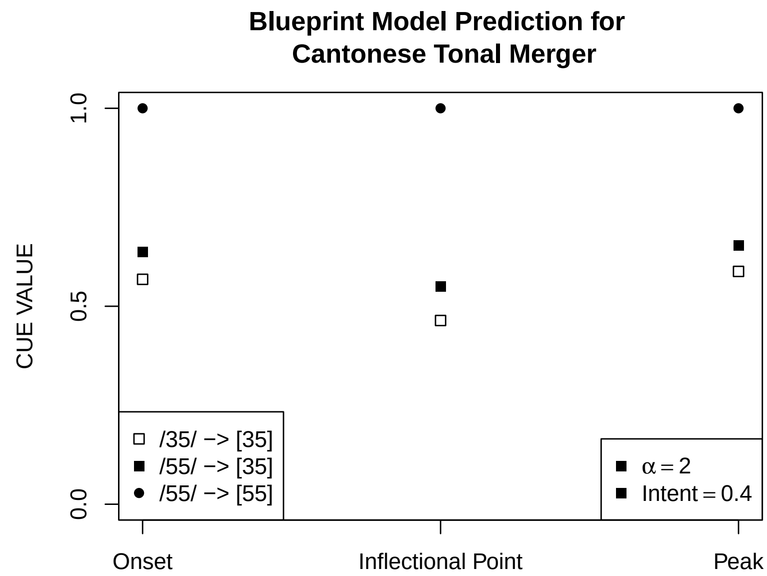


Figure 4: Simulated cue values for Yu (2007) tone-sandhi data

form and the actual surface form.

## 4.6 Epenthesis in Arabic

Lebanese Arabic speakers epenthesize an [i] vowel to break up word final CC clusters. Gouskova and 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 final 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 [ɨ] or [ə] – and, arguably, to zero" (emphasis original). While they use Optimality Theory with Candidate Chains to explain these findings (OT-CC; McCarthy, 2007), 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 epenthesized 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 0ms when *intent* was 1 to the target duration for an [i] vowel when *intent* was 0. If the level of *intent* is in between 0 and

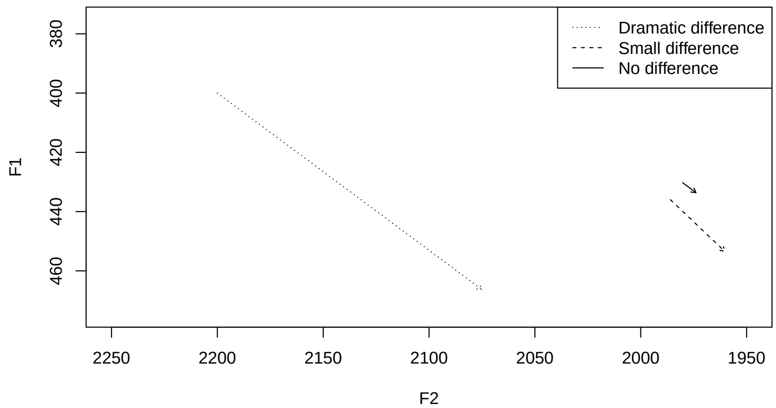


Figure 5: Simulated formant values for lexical and epenthetic [i] vowels based on Hall (2013) for a dramatic difference speaker, a small difference speaker, and a no difference speaker

1 then the duration of the epenthetic vowel will always be closer to zero, which is exactly what Gouskova and Hall (2009) find.

Hall (2013) follows up on Gouskova and 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 (61ms) and epenthetic (60ms) vowels at the group level.[11] Hall (2013) hypothesizes 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 and 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) and the other two speakers were each assigned the vector (450,2000) to indicate a more central vowel.

[11] Individual differences were not reported.

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 the Hall (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 emphasize 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 regards to phonetic representations. If phonetic representations are acoustic targets, then a zero morpheme would have to have some type of acoustic target even if its duration was also 0. One plausible set of values is those corresponding to the default/neutral segment within the language (Archangeli, 1984; Broselow, 1984; Pulleyblank, 1988; McCarthy and Prince, 1994). In the simulation above, we chose a neutral vowel (schwa) as 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 and Hall (2009) claim that the phenomenon at hand is a case of incomplete neutralization, 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 of a given token plays no role in the synchronic phonetic realization. This is what allows for a unified explanation of the 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 towards a different scenario: homophones. It has been reported that many homophonic pairs have subtle phonetic differences, most notably along the temporal dimension (Walsh and Parker, 1983; Losiewicz, 1995; Gahl, 2008; Lohmann, 2018a,b). Like neutralized pairs, homophones share the same surface phonological form, but unlike neutralized 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 realization of phonological units (Fosler-Lussier and Morgan, 1998; Bybee, 2001; Jurafsky et al., 2001; Bell et al., 2009). Leslau (1969) reports that the Arab Grammarians were tuned in 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 organization of the grammar.

## 5.2 Adding Frequency to the Blueprint Model of Production

In the same way that Intent is an input to the Phonetics function in 4.2, frequency information is yet another input. Frequency is represented as a function $F$ and the phonetics function is updated accordingly: $A_{\text{BP}} :: L \to P \to I \to F \to \{PR\}$. In other words, the 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 $F :: L \to \mathbb{R}$. Since the Lexicon 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 the 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 phonetic representation. 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) $\tau = \pi(L) \times I^{\alpha} + \pi(P(L)) \times (1 - I^{\alpha}).$

Recall that $L$ contains URs and $P(L)$ returns SRs. 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 sub-programs within the larger phonetics function $A_{\text{BP}}$.

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 phonetic representation 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 place them directly in the lexicon or the phonological grammar. Instead, they are just one more factor alongside the lexicon, phonological grammar, and pragmatic intent that influences production.

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 and 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 involve temporal cues, our implementation would equally apply to spectral cues as well. This raises the question as to whether or not frequency information can also influence non-temporal cues. The answer appears to be yes. In a recent review of phonetic reduction, Clopper and Turnbull (2018) discuss ways in which 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 and Solomon, 2004). Crucially, Munson and 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.

## 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, pp. 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 homophonous pair 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.[12]

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

[12]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 adjective. A closer analysis may show that splitting these forms apart may show duration and frequency values that do follow the general trend. This is beyond the current scope of the paper.

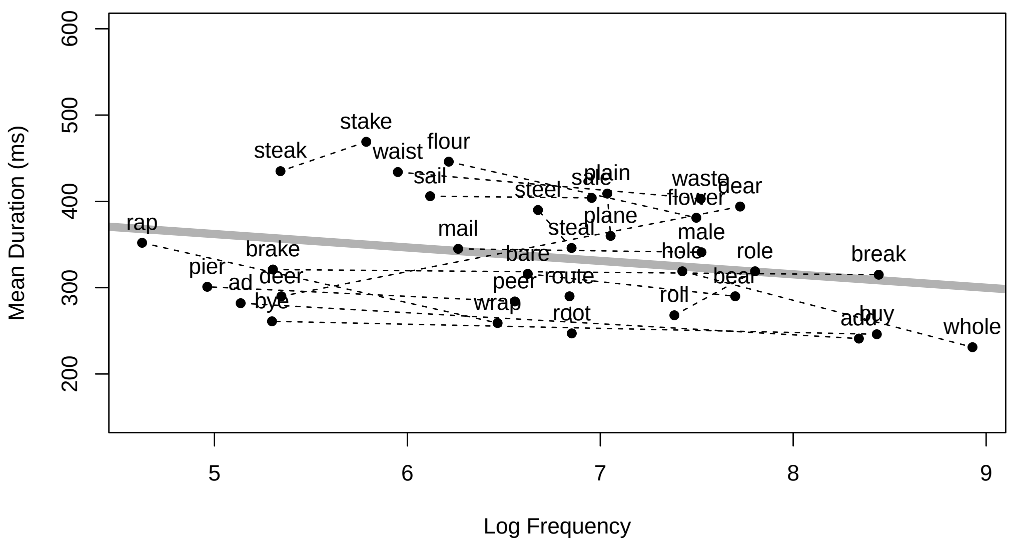


Figure 6: Average duration and log frequency for 17 words and their homophone twin. These data come from the Switchboard corpus (Godfrey et al., 1992). Dashed thin lines connect all homophonous pairs. The thick gray line is the output of a linear model ($\mathcal{I}$) of these points showing a general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

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 $\beta_1 \times LogFreq(x)$ removed. In other words, the intercept for each phonological form is the hypothesized target value.[13]

To relate $\mathcal{S}$ to the duration scaling function $\delta(F(L))$ above, it is necessary to do some rearranging of terms. In its current form, $\mathcal{S}$ is similar to the Pierrehumbert (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 $y = \beta_1 \times LogFreq(x) + k$. Basic algebra derives an equivalent form: $y = k \times (1 + \frac{\beta_1 \times LogFreq(x)}{k})$. Since $\beta_1$, the slope coefficient, is negative and $LogFreq(x)$ is guaranteed to be non-negative, the value of $(1 + \frac{\beta_1 \times LogFreq(x)}{k})$ is guaranteed to be less than 1. As long as $\beta_1 \times LogFreq(x)$ is less than or equal to $k$, the value of $(1 + \frac{\beta_1 \times LogFreq(x)}{k})$ is also guaranteed

[13] A reviewer points out that since the $\beta$ weight on frequency is a free parameter that is fit to the data, then 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 the introduction, the BMP, as a computational level description, has nothing to say about this issue and freely overgenerates. Other kinds of evidence or principals will be necessary to constrain it.

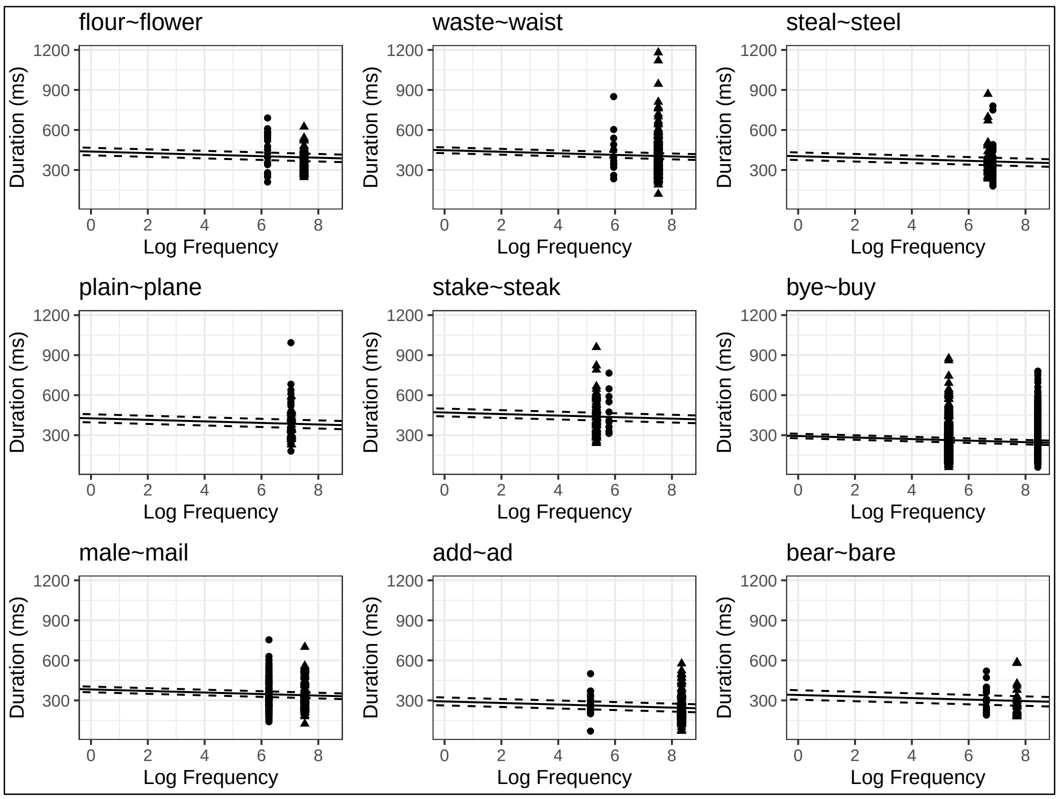


Figure 7: Frequency and Duration information for individual tokens of 9 randomly selected homophonous pairs. Each plot represents a single pair. The solid black lines are the predicted linear relationship for that phonological form. Dashed lines indicate 95% confidence intervals.

to be greater than or equal to 0. Under these conditions, this works exactly as a scaling factor in 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)
$$1 + \frac{\beta_1 \times LogFreq(x)}{\beta_0 + \sum_{i=1}^{|L|} \beta_i \times [l_i = x]}$$

Figure 7 shows the individual duration values for nine randomly selected homophonous pairs as well as the output of $\mathcal{S}$ for each phonological form. $\mathcal{S}$ has a significant effect for frequency ($\beta$ = -5.761, t=-3.561, p < 0.001). To illustrate how this works, consider the pair bye$\sim$buy. $\mathcal{S}$ predicts an intercept of 293.852 for this phonological form and therefore provides the equation $\hat{y} = 293.852 - 5.761 \times LogFreq(x)$. 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*, the 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.

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 focus 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 phonetic representation; 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 paper introduced an abstract model of language production called the BLUEPRINT MODEL OF PRODUCTION which is characterized 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 neutralization, 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 $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$. As discussed in section 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 each argument can be given one at 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. $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$
b. $A^{l}_{\text{BP}} :: P \rightarrow I \rightarrow F \rightarrow \{PR\}$
c. $A^{l,p}_{\text{BP}} :: I \rightarrow F \rightarrow \{PR\}$
d. $A^{l,p,i}_{\text{BP}} :: F \rightarrow \{PR\}$

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, $A^{'}_{\text{BP}} :: (L, P) \rightarrow (I, F) \rightarrow \{PR\}$. 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 paper, it is possible to switch $(I, F)$ to a cover type $E$ which stands in for all of the information other than the lexicon and phonology that go into the production process. With this in mind, it is possible to have a partially-saturated function with type $A'^{l,p}_{BP} :: E \rightarrow \{PR\}$. 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 phonetics 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 it is 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 paper careful phonetic experimentation which has yielded evidence for the importance of extra-grammatical factors on production. Additionally, other experimental work has shown the importance of maintaining categorical phonological knowledge (Du and Durvasula, in press; Mai et al., 2022). Considering (van Rooij and Baggio, 2020)'s characterization 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 paper 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 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 externalize 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.[14] 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

[14]We recognize that certain processes are optional and/or gradient, but would argue that phonological accounts of them still automatically takes place. In other words, the optionality and gradience is determined by the automatic application of the phonology function.

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 paper is that phenomena like incomplete neutralization 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).

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 (c.f. 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 formalized version of what Du and Durvasula (in press) 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, p. 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 paper, along with simulations in our case studies, provide multiple entry points for further investigation of the BMP on its own terms or in comparison to other models of the interface. We began some comparison with the research using dyanmical systems because of its significant influence on our thinking. A reviewer points out that Jurafsky et al. (2002, figure 3) and Shaw and 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 realizations of other kinds of phenomena including deletion, the realization of absolutely neutralized segments, morphological boundary effects, and optionality.

In this paper, we formalize the BMP using typed functions and show how the BMP architecture allows for the simulation of systematic phonetic gradience found in incomplete neutralization, 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 neutralization 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.

# References


Albright, A. and Hayes, B. (2006). Modelling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Fanselow, G., Féry, C., Schlesewsky, M., and Vogel, editors, *Gradience in Grammar*. Oxford University Press.

Archangeli, D. B. (1984). *Underspecification in Yawelmani phonology and morphology*. PhD thesis, Massachusetts Institute of Technology.

Avery, P. and Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In Hall, T. A., editor, *Distinctive feature theory*, pages 41–70. de Gruyter, Berlin.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1996). The celex lexical database (cd-rom).

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, 60(1):92–111.

Bermúdez-Otero, R. (2007). Diachronic phonology. In de Lacy, P., editor, *The Cambridge handbook of phonology*, pages 497–517. Cambridge.

Bloomfield, L. (1933). *Language history: from Language (1933 ed.).* Holt, Rinehart and Winston.

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology*, 36(1):1–36.

Braver, A. and Kawahara, S. (2016). Incomplete neutralization in Japanese monomoraic lengthening. In *Proceedings of the Annual Meetings on Phonology*, volume 2.

Broselow, E. (1984). Default consonants in amharic morphology. *MITWPL*, 7:15–31.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Chandlee, J. (2014). *Strictly local phonological processes*. PhD thesis, University of Delaware.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English.* ERIC.

Church, A. (1932). A set of postulates for the foundation of logic. *Annals of mathematics*, pages 346–366.

Church, A. (1933). A set of postulates for the foundation of logic. *Annals of mathematics*, pages 839–864.

Clopper, C. G. and Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., and Zellers, M., editors, *Rethinking reduction: interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, pages 25–72. Mouton de Gruyter.

Coats, H. S. and Harshenin, A. P. (1971). On the phonological properties of Russian. *The Slavic and East European Journal*, 15(4):466–478.

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language & Linguistic Theory*, 26:289–337.

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology*, 10(1):43–81.

Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory*, 16:1–31.

Coleman, J. and Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*.

Cooper, R. P. and Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, 27:42–49.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press.

Curry, H. B. (1980). Some philosophical aspects of combinatory logic. In Barwise, J., Keisler, H. J., and Kunen, K., editors, *The Kleene Symposium*, volume 101 of *Studies in Logic and the Foundations of Mathematics*, pages 85–101. Elsevier.

Deshmukh, N., Ganapathiraju, A., Gleeson, A., Hamaker, J., and Picone, J. (1998). Resegmentation of switchboard. In *ICSLP*. Syndey.

Dinnsen, D. and Garcia-Zamor, M. (1971). The three degrees of vowel length in German. *Research on Language & Social Interaction*, 4(1):111–126.

Dinnsen, D. A. and Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of phonetics*, 12(1):49–60.

Dmitrieva, O., Jongman, A., and Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of phonetics*, 38(3):483–492.

Du, N. and Durvasula, K. (in press). Phonetically incomplete neutralization can be phonologically complete: Evidence from Huai'an Mandarin. *Phonology*.

Duhem, P. M. M. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press.

Ernestus, M. (2011). Gradience and categoricality in phonological theory. In *The Blackwell companion to phonology*, pages 2115–2136. Wiley-Blackwell.

Ernestus, M. T. C. and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79(1):5–38.

Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive science*, 6(3):205–254.

Feyerabend, P. (1965). Reply to criticism: comments on smart, sellars and putnam. In Cohen, R. and Wartofsky, M., editors, *Proceedings of the Boston Colloquium for the Philosophy of Science II*, pages 223–261. Springer.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, 18(1):7–44.

Fodor, J. A. (1974). Special sciences, (or: the disunity of science as a working hypothesis). *Synthese*, 28:97–115.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Fosler-Lussier, E. and Morgan, N. (1998). Effects of speaking rate and word frequency on conversational pronunciations. In *Modeling Pronunciation Variation for Automatic Speech Recognition*.

Fougeron, C. and Steriade, D. (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? In *Fifth European Conference on Speech Communication and Technology*.

Frixione, M. (2001). Tractable competence. *Minds and Machines*, 11:379–397.

Gafos, A. I. (2006). Dynamics in grammar: Comment on Ladd and Ernestus & Baayen. *Laboratory phonology*, 8(4):51.

Gafos, A. I. and Benus, S. (2006). Dynamics of phonological cognition. *Cognitive science*, 30(5):905–943.

Gafos, A. I., Charlow, S., Shaw, J. A., and Hoole, P. (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics*, 44:152–166.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

Gerfen, C. (2002). Andalusian codas. *Probus*, 14(2):247–277.

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, volume 1, pages 517–520. IEEE Computer Society.

Goldrick, M. (2000). Turbid output representations and the unity of opacity. In *North East Linguistics Society*, volume 30.

Gordon, M. (2004). *Syllable weight*, chapter 9, pages 277–312. Cambridge University Press Cambridge.

Gorman, K. and Sproat, R. (2021). *Finite-State Text Processing*. Morgan & Claypool Publishers.

Gouskova, M. and Hall, N. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. *Phonological argumentation: Essays on evidence and motivation*, pages 203–225.

Guest, O. and Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior*, pages 1–15.

Gussmann, E. (2007). *The phonology of Polish*. Oxford University Press.

Hale, M. and Reiss, C. (2000). "Substance abuse" and "dysfunctionalism": current trends in phonology. *Linguistic inquiry*, 31(1):157–169.

Hale, M. and Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.

Hall, N. (2013). Acoustic differences between lexical and epenthetic vowels in lebanese arabic. *Journal of Phonetics*, 41(2):133–143.

Halle, M. and Stevens, K. (1971). A note on laryngeal features. *MIT Res. Lab. Electron. Q. Prog. Rep.*, 101:198–213.

Hayes, B., Kirchner, R., and Steriade, D. (2004). *Phonetically Based Phonology*. Cambridge University Press.

Heinz, J. (2018). The computational nature of phonological generalizations. In Hyman, L. and Plank, F., editors, *Phonological Typology*, Phonetics and Phonology, chapter 5, pages 126–195. De Gruyter Mouton.

Hinton, G. E. and Anderson, J. A., editors (1981). *Parallel models of associative memory: updated edition*. Erlbaum.

Iverson, G. K. and Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12(3):369–396.

Jassem, W. and Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, 17(4):317–325.

Jurafsky, D., Bell, A., and Girand (2002). The role of the lemma in form variation. In Gussenhoven, C. and Warner, N., editors, *Laboratory phonology VII*, pages 1–34.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. phonetic linguistics: Essays in honor of peter ladefoged, ed. victoria a. fromkin.

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

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of phonetics*, 18(3):321–334.

Kingston, J. (2019). The interface between phonetics and phonology. *The Routledge handbook of phonetics*, pages 359–400.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3):419–454.

Kopkalli, H. (1994). *A phonetic and phonological analysis of final devoicing in Turkish*. PhD thesis, University of Michigan.

Labov, W., Yaeger, M., and Steiner, R. (1972). *A quantitative study of sound change in progress*, volume 1. US Regional Survey.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, volume 4, page 91–196. Cambridge University Press.

Lambert, D., Rawski, J., and Heinz, J. (2021). Typology emerges from simplicity in representations and learning. *Journal of Language Modelling*, 9.

Leslau, W. (1969). Frequency as determinant of linguistic changes in the Ethiopian languages. *Word*, 25(1-3):180–189.

Lionnet, F. (2017). A theory of subfeatural representations: the case of rounding harmony in Laal. *Phonology*, 34(3):523–564.

Lohmann, A. (2018a). Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs. *Journal of Linguistics*, 54(4):753–777.

Lohmann, A. (2018b). Time and thyme are not homophones: A closer look at gahl's work on the lemma-frequency effect, including a reanalysis. *Language*, 94(2):e180–e190.

Losiewicz, B. (1995). Word frequency effects on the acoustic duration of morphemes. *Journal of the Acoustic Society of America*, 97:32–43.

Łukaszewicz, B. (2021). The dynamical landscape: phonological acquisition and the phonology–phonetics link. *Phonology*, 38(1):81–121.

Mai, A., Riès, S., Ben-Haim, S., Shih, J., and Gentner, T. (2022). Phonological contrasts are maintained despite neutralization: an intracranial eeg study. In *Proceedings of the Annual Meetings on Phonology*, volume 9.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. CUMINCAD.

McCarthy, J. J. (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Equinox Publishing (Indonesia).

McCarthy, J. J. and Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In *North East Linguistics Society*, volume 24.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological science*, 2(6):387–395.

Mohanan, K. P. (1986). *The theory of lexical phonology*. Springer.

Munson, B. and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, 47(5):1048–1058.

Myers, S. (2000). Boundary disputes: The distinction between phonetic and phonological sound patterns. *Phonological knowledge: Conceptual and empirical issues*, 245:272.

Newell, A. and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*, 65(3):151.

Nicenboim, B., Roettger, T. B., and Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in german. *Journal of Phonetics*, 70:39–55.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech*, pages 189–216. Springer.

Ohala, J. J. (1990). There is no interface between phonology and phonetics: a personal view. *Journal of phonetics*, 18(2):153–172.

Ohala, J. J. (1992). The costs and benefits of phonological analysis. *The linguistics of literacy*, pages 211–238.

Pierce, B. C. (2002). *Types and programming languages*. MIT press.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of phonetics*, 18(3):375–394.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*, 7.

Popper, K. (1959). *The logic of scientific discovery*. Routledge.

Port, R. and Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of phonetics*, 17:257–282.

Port, R., Mitleb, F., and O'Dell, M. (1981). Neutralization of obstruent voicing in German is incomplete. *The journal of the Acoustical Society of America*, 70(S1):S13–S13.

Port, R. F. and Leary, A. P. (2005). Against formal phonology. *Language*, 81(4):927–964.

Port, R. F. and O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of phonetics*.

Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in phonology*.

Priva, U. C. and Gleason, E. (2020). The causal structure of lenition: A case for the causal precedence of durational shortening. *Language*, 96(2):413–448.

Pulleyblank, D. (1988). Underspecification, the feature hierarchy and Tiv vowels. *Phonology*, 5(2):299–326.

Putnam, H. (1967). Psychological predicates. In Capitan, W. H. and Merill, D. D., editors, *Art, mind, and religion*, pages 37–48. University of Pittsburgh Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1):20–43.

Ramer, A. M. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics*, 4(24):477–489.

Reiss, C. (2018). Substance free phonology. *The Routledge handbook of phonological theory*, pages 425–452.

Reiss, C. and Volenec, V. (2020). Conquer primal fear: Phonological features are innate and substance-free. https://ling.auf.net/lingbuzz/005683.

Revithiadou, A. (2008). Colored turbid accents and containment: a case study from lexical stress. In Blaho, S., Bye, P., and Krämer, M., editors, *Freedom of Analysis?* De Gruyter Mouton.

Roark, B. and Sproat, R. (2007). *Computational approaches to morphology and syntax*, volume 4. Oxford University Press.

Roon, K. D. and Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, 89:222–243.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1988). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. MIT press.

Saltzman, E. L. and Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4):333–382.

Shaw, J. and Tang, K. (2023). A dynamic neural field model of leaky prosody: proof of concept. In *Proceedings of the Annual Meetings on Phonology*, volume 10.

Slowiaczek, L. M. and Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of phonetics*.

Smolensky, P. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. *Ms. Johns Hopkins University and Northwestern University. Available as ROA*, 1286.

Solé, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and speech*, 35(1-2):29–43.

Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1):1–23.

Solé, M.-J. (2007). Controlled and mechanical properties in speech. *Experimental approaches to phonology*, pages 302–321.

Strycharczuk, P. (2019). Phonetic detail and phonetic gradience in morphological processes.

Tuller, B., Case, P., Ding, M., and Kelso, J. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human perception and performance*, 20(1):3.

Van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua*, 118(9):1362–1374.

Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive science*, 32(6):939–984.

van Rooij, I. and Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*.

Van Rooy, B., Wissing, D., and Paschall, D. D. (2003). Demystifying incomplete neutralisation during final devoicing. *Southern African Linguistics and Applied Language Studies*, 21(1-2):49–66.

Volenec, V. and Reiss, C. (2017). Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics*, 11:251–294.

Walsh, T. and Parker, F. (1983). The duration of morphemic and nonmorphemic /s/ in English. *Journal of Phonetics*, 11:201–206.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of phonetics*, 32(2):251–276.

Westbury, J. R. and Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of linguistics*, pages 145–166.

Wheeler, M. W. (2005). *The phonology of Catalan*. Oxford University Press on Demand.

Winter, B. and Röettger, T. (2011). The nature of incomplete neutralization in German: Implications for laboratory phonology. *Grazer Linguistische Studien*, 76:55–74.

Yang, C. (2016). *The Price of Productivity*. MIT Press.

Yu, A. C. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 24(1):187–214.

Yu, A. C. (2011). Mergers and neutralization. *The Blackwell companion to phonology*, pages 1–27.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. *Phonetically based phonology*, pages 157–190.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of phonetics*, 28(1):69–102.