The blueprint model of production

Scott Nelson and Jeffrey Heinz

April 23, 2023

Abstract

This paper introduces an abstract characterization of the phonetics-phonology interface called the blueprint model of production. The blueprint model of production is formalized using typed functions. The standard modular feed-forward view is that the phonetic form of a lexical item is the output of the composition of a phonetic function f and a phonological function g: $f(g(x))$. The central idea of the blueprint model of production is that the phonetic form of a lexical item is instead the output of a higher order function that takes the phonology function as one of its inputs: $f(g(), x)$. We show how this formalization of the production process is able to account for phenomena such as incomplete neutralization, some cases of near merger, and variation in homophone duration while maintaining a discrete phonological grammar. Consequently, these type of systematic fine-grained phonetic patterns do not necessarily provide evidence against formal phonology.

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.

Under these types of architectural commitments, 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, 1992; Pierrehumbert, 2002; Port and Leary, 2005). This often results in researchers reconceptualizing the individual modules by making the phonology 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).

This paper takes a different approach. Rather than proposing changes to the modules directly, we provide an alternate architecture called the BLUEPRINT MODEL OF PRODUCTION 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 BLUEPRINT MODEL OF PRODUCTION 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 as one of its inputs: $f(g(), x)$.

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. With an architecture such as the BLUEPRINT MODEL OF PRODUCTION, problematic phenomena such as incomplete neutralization and homophone duration are accounted for in a straightforward manner. Furthermore, it is shown that these date can be accounted for while maintaining a discrete, symbolic phonological grammar. This point is made not to say that it has been proven that phonology must be discrete, but rather that phonology does not necessarily need to be gradient. Simulations show that this type of architecture allows for a discrete phonology and the ability to account for well documented phonetic detail. These simulations are presented with a hope that the architectural claims made in this paper will help clarify what types of facts need to be accounted for by the phonological grammar proper.

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 BLUEPRINT MODEL OF PRODUCTION is couched.

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 BLUEPRINT MODEL OF PRODUCTION using typed functions. The next two sections provide several case studies which show how an instantiation of the BLUEPRINT MODEL OF PRODUCTION 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 homo-

[1] For the remainder of this paper these terms are used interchangeably.

phone durations are discussed under the purview of the BLUEPRINT MODEL OF PRODUCTION. 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 this 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 model 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 corse-grained phonological assimilation (and, of course, 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 phonological primes 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 modeled 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 do 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.

If we reject the previous two accounts in favor of separate phonology and phonetics 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. 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 is in relation to representations (Smolensky and Goldrick, 2016; Lionnet, 2017). A third way is grammatical acceptability judgements (Coleman and Pierrehumbert, 1997; Coetzee and Pater, 2008). 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.

On the other hand, 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 τ_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 that there are a wide-range of views to consider when discussing a phonetically based phonology. On one end you have phonetics *in* phonology while the other end 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 contrast, the substance free phonology framework (Hale and Reiss, 2000, 2008; Reiss, 2018) sharply demarcates the boundaries 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 string based on 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 while using a constraint grammar implemented with dynamical systems.

Instead, we 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).

Crucial to the BLUEPRINT MODEL OF PRODUCTION 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 BLUEPRINT MODEL OF PRODUCTION 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 BLUEPRINT MODEL OF PRODUCTION 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 BLUEPRINT MODEL OF PRODUCTION.

3 The Blueprint Model of Production

This section presents the BLUEPRINT MODEL OF PRODUCTION in detail. We will begin with a discussion of the production process within generative phonology and then transition into a formal explanation of the proposal. The BLUEPRINT MODEL OF PRODUCTION 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 BLUEPRINT MODEL OF PRODUCTION just as there are many ways to *instantiate* the phonetics-phonology interface within a feed-forward model of the interface.

There are two essential points to understanding the BLUEPRINT MODEL OF PRODUCTION. First, it concretely models the production process with multiple, simultaneous factors, of which phonology is just one.[2]Second, the whole phonological module is a factor in production, not just the representations it outputs. Gafos and Benus (2006) express a similar idea. Their concluding sentence explains that their model "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, their model provides a specific instantiation of the BLUEPRINT MODEL OF PRODUCTION's architecture, but it is not the only one possible.

In this context, our contributions are as follows. First, we show how the BLUEPRINT MODEL

[2] 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.

OF PRODUCTION reconceptualizes the relationship between phonology and phonetics. One outcome of this is that the BLUEPRINT MODEL OF PRODUCTION 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. Finally, we are able to situate the BLUEPRINT MODEL OF PRODUCTION to previous work on the phonetics-phonology interface using a type-functional analysis. In particular, we show exactly how the BLUEPRINT MODEL OF PRODUCTION 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. 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 set of instructions 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 how to assemble materials into forms, 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. Under this architecture, the lexical form is not invisible to the Phonetics Station (cf. Van Oostendorp (2008)).

Phonetic studies over the past 40 years have lead some researchers to question the ability of formal phonology to accurately portray the production process (Bybee, 1999; Port and Leary, 2005). We agree that language production is a process that involves multiple interacting factors and therefore cannot *solely* be understood in terms of the representation output by a single symbolic computational system. We disagree that there is a need to abandon, or even diminish, the role of a formal phonology in the production process. The BLUEPRINT MODEL OF PRODUCTION provides an alternative formulation of the production process that accounts for fine-grained sub-symbolic phonetic detail while simultaneously permitting a discrete phonology.

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 we might think of the phonology module as a function that maps a UR to an SR and the phonetics module as a function that maps an SR to a PR (Phonetic Representation). The BLUEPRINT MODEL OF PRODUCTION continues to view the phonology module as a function that maps a UR to an SR but models the phonetics module as a higher order function that *takes the phonology module function as an input.* Instead of a single UR, the entire lexicon is also an input to the phonetics module function in order to generalize over all lexical items. The phonetics module is therefore a function with two inputs: the lexicon and the phonology module function; 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 BLUEPRINT MODEL OF PRODUCTION.

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 BLUEPRINT MODEL OF PRODUCTION can be related directly to the feed-forward model via function application. We also show that the BLUEPRINT MODEL OF PRODUCTION 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

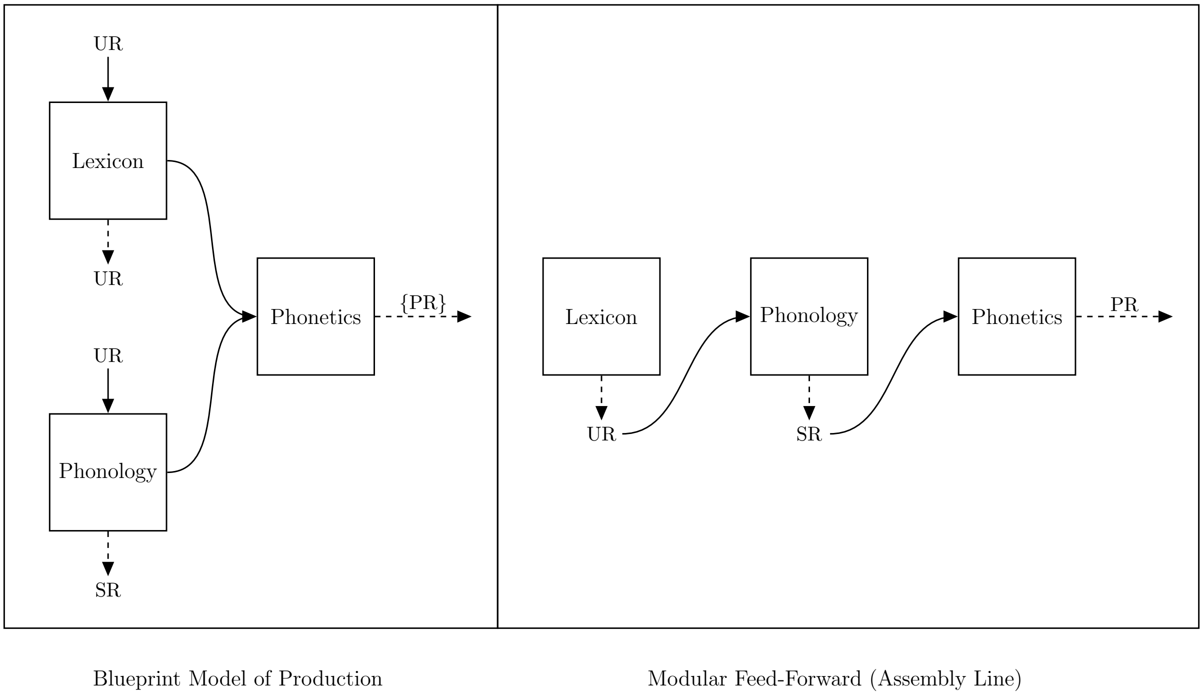

Figure 1: Visual comparison of the architecture for modular feed-forward models and the BLUEPRINT MODEL OF PRODUCTION. Each box represents a function/module. Solid lines represent the inputs to each function while dashed lines represent the outputs of each function.

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 would be equal to, "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 numbers to a single real number.

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 $f(x)$ and X, where $f(x)$ is a function of type $f :: X \to Y$ that takes things of type x as its input, and $\vec{x}$ is an array of length n that contains x's, $map(f(x), \vec{x})$ will apply function $f(x)$ to every individual element of $x \in \vec{x}$ and return the array $[f(x_0), \ldots, f(x_n)]$. To give a concrete example, consider the function $+_1(x) = x + 1$ 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 $map(+_1(), [-23, 1, 9, 307]) = [-22, 2, 10, 308]$. The map function is not limited to numerical data types/functions and works just as well over strings. If we assume a string abc is equal to $[a, b, c]$ then we could also use the map function to manipulate strings. Suppose we had a function where $g(a) = a$ and $g(b) = g(c) = b$ then $map(g, abc) = abb$. To summarize, the function type of $map()$ is given by $map :: (X \to Y) \to [X] \to [Y]$.

We now move to a discussion of function application. Function application is the process 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 $X \to Y$ (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 $function\text{-}application :: X \to (X \to Y) \to Y$. The notion of function application is important for our analysis because it allows us to relate the BLUEPRINT MODEL OF PRODUCTION 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.[3]

Name	Meaning	Type
L	Lexicon	$UR \to UR$
P	Phonology	$UR \to SR$
A_{MFF}	Phonetics$_{\text{MFF}}$	$SR \to PR$
A_{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 BLUEPRINT

[3] In Figure 1 the Lexicon has type $UR \to UR$. This is typically implemented as the identity function (Roark and Sproat, 2007). This is an abstraction to facilitate the analysis.

MODEL OF PRODUCTION. To start, the phonetics module in the modular feed-forward model has the following type.

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

This idealizes the phonetics module as a map from surface representations to phonetic representations. The definition of function application from above can then be used to decompose SR into $UR \rightarrow (UR \rightarrow SR)$.

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

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

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

To complete this reconceptualization we change UR to L in order to generalize over the entire lexicon (Roark and Sproat, 2007). 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) $A_{\text{BP}} :: L \rightarrow P \rightarrow \{PR\}$

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 BLUEPRINT MODEL OF PRODUCTION and the modular feed-forward model an underlying form is ultimately transformed into a phonetic representation. The main difference is phonology is no longer 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 BLUEPRINT MODEL OF PRODUCTION, 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 BLUEPRINT MODEL OF PRODUCTION 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. Under the BLUEPRINT MODEL OF PRODUCTION, this is accomplished by adding the intent (I) as one of the arguments to production: $A :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$. Claims of how exactly the phonetics module (A) integrates the lexicon (L), the phonological module (P), and the intent (I) should be recognized as claims of how the BLUEPRINT MODEL OF PRODUCTION can be instantiated in various particular ways. As with the feed-forward model, the BLUEPRINT MODEL OF PRODUCTION itself, as an abstract characterization of the phonetics-phonology interface, has little to say about specific instantiations.

3.3 Currying and Uncurrying

This section relates the BLUEPRINT MODEL OF PRODUCTION 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* and *derived* word-final, voiceless obstruents (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 BLUEPRINT MODEL OF PRODUCTION. To explain, consider the fact that since functions can return functions in general, functions with multiple arguments do not need to saturate them all at once. If fewer than the totality of arguments is given, then a *function* can be returned.

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

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

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

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

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

Combining (6) and (3) reveals that the BLUEPRINT MODEL OF PRODUCTION can be characterized as shown below.

(7) $A :: UR \rightarrow P \rightarrow PR = UR \rightarrow curry(A')$

Generalizing over the lexicon again, we get the same type for the BLUEPRINT MODEL OF PRODUCTION.

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

This shows precisely the relation between containment theories of phonology under the feed-forward model and *any* theory of phonology computing functions $UR \rightarrow SR$ with the BLUEPRINT MODEL OF PRODUCTION. It also highlights the essential difference between the BLUEPRINT MODEL OF PRODUCTION 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. We argue that these disparate phenomena can be unified under the BLUEPRINT MODEL OF PRODUCTION approach to the phonetics-phonology interface. Furthermore, we argue this architecture provides a way to move beyond the "abstract" vs. "episodic" debate Pierrehumbert (2016) within *phonology*. It allows for clear interpretation of how individual factors interact during the global production process which helps clarify what exactly the phonological grammar needs to account for and what phonetic factors can be explained through extra-grammatical means. Incomplete neutralization (Port et al., 1981; Port and O'Dell, 1985) and homophone durational variation (Gahl, 2008) are two phenomena we think best highlight the strength of the BLUEPRINT MODEL OF PRODUCTION.

4 Incomplete Neutralization

This section first provides background on incomplete neutralization. After the empirical facts have been laid out, we discuss how the BLUEPRINT MODEL OF PRODUCTION 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.[4]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).

(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

[4] 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).

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 as well (Catalan (Dinnsen and Charles-Luce, 1984); Dutch (Warner et al., 2004); Polish (Slowiaczek and Dinnsen, 1985); Russian (Dmitrieva et al., 2010); 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.

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 model of 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 Adding Intent to the BLUEPRINT MODEL OF PRODUCTION

Section 3.2 provided a formal characterization of the production process. The focus in that section was to show how the BLUEPRINT MODEL OF PRODUCTION 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, the current formulation of the BLUEPRINT MODEL OF PRODUCTION lacks an explicit parameter for controlling the speakers intent. The BLUEPRINT MODEL OF PRODUCTION can be extended in a way that allows for the scaling of UR and SR information by including an intent variable in the input. We use I for the intent variable, updating our function to be: $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$.

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. However, it is beneficial to provide one possible concrete instantiation to show how the BLUEPRINT MODEL OF PRODUCTION can model the gradient incomplete neutralization data while maintaining a discrete phonology. This instantiation is used throughout 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 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. 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.[5]

Since the BLUEPRINT MODEL OF PRODUCTION 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.

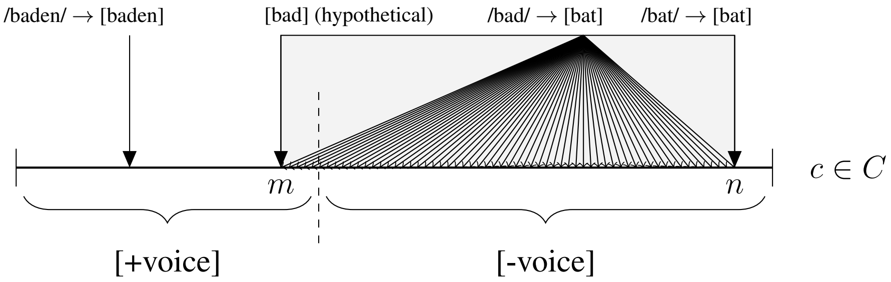

Figure 2: Hypothetical Cue space

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_{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

[5] 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.

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 α 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 model 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. If you look at the line for $\alpha = 20$, it is only when intent is greater than 0.8 that it raises above what is effectively 0. 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. 5).

We would like to end this section by once again stressing that while the above method is one way in which the BLUEPRINT MODEL OF PRODUCTION can be implemented to account for incomplete neutralization, it is not the only way. As discussed above, the production model of Gafos and Benus (2006) has been rather influential on our formulation here. 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 final system is a linear combination of these two equations where the equation for the faithfulness constraint contains a control parameter θ that corresponds to intent and allows the faithfulness constraint (i.e. - underlying form) to have more or less influence on the final output. We therefore see this nonlinear dynamic model as another possible instantiation of the BLUEPRINT MODEL OF PRODUCTION in the same way that Pierrehumbert (2002) described essentially a class of production models when using the term modular feed-forward.[6]

In the remaining parts of this section we present three case studies to show how our instantiation of the BLUEPRINT MODEL OF PRODUCTION 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.

4.3 Final Devoicing in German

The intent argument was added to the BLUEPRINT MODEL OF PRODUCTION 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 α parameter can interact to model these findings. Specifically, their study found that only certain cues were acoustically incomplete. Burst duration was the main cue found to be incomplete, while preceding vowel duration, a cue that has been found to be robust for German voicing contrast (Chen, 1970; Braunschweiler, 1997), was complete. Our model accounts for this distinction.

Mean values for both vowel duration and burst 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). 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,

[6] Furthermore, Gafos and Benus (2006, fn. 4) suggest that their model and one similar to what we described above based on directly scaling UR/SR information would essentially be extensionally equivalent. One difference of course being that our suggestion still remains more abstract which allows for it to be adapted by a wider base of phonologists without adopting nonlinear dynamics. Gafos and Benus (2006) do make the additional claim that only the nonlinear dynamic model can account for their Hungarian vowel harmony data. We disagree and believe that the phonetic results they present are also consistent with a UR/SR scaling account if we assume an absolute neutralization account of Hungarian vowel harmony (Vago, 1976).

Condition		Vowel Duration (Mean)	Ratio	Burst Duration (Mean)	Ratio
1A	/d/	138.19	0.95	20.08	0.78
	/t/	145.67		25.59	
1B	/d/	135.03	0.96	16.54	0.58
	/t/	140.94		28.35	
2	/d/	178.14	1.02	32.87	0.83
	/t/	174.80		39.37	
3	/d/	203.34	1.00	25.20	0.29
	/t/	203.57		85.63	
4	/d/	235.83	0.97	59.06	0.89
	/t/	243.06		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/.

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 allows us to abstract away from the condition specific variation in duration. The ratios reported in Table 2 can therefore be used to estimate intent values.

Each subfigure within 3 shows the estimated cue values for both burst duration and vowel duration. Since burst duration (represented as $+$) was found to significantly vary between derived and faithful surface /t/ segments in the pooled data, but vowel duration (represented as $\times$) was not, the α parameter was set to 1 for burst duration and 20 for vowel duration. In general, the ratio of vowel duration for underlying /d/ segments to underlying /t/ segments was 0.95 or higher for each condition. 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.[7]

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 vowel duration. Within the left plot, the cue values are shown to vary between burst duration ($+$) and proceeding vowel 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

[7] In their discriminant analysis of their perceptual experiment, 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. Better understanding the interplay of perception and production under the BLUEPRINT MODEL OF PRODUCTION is left for future work.

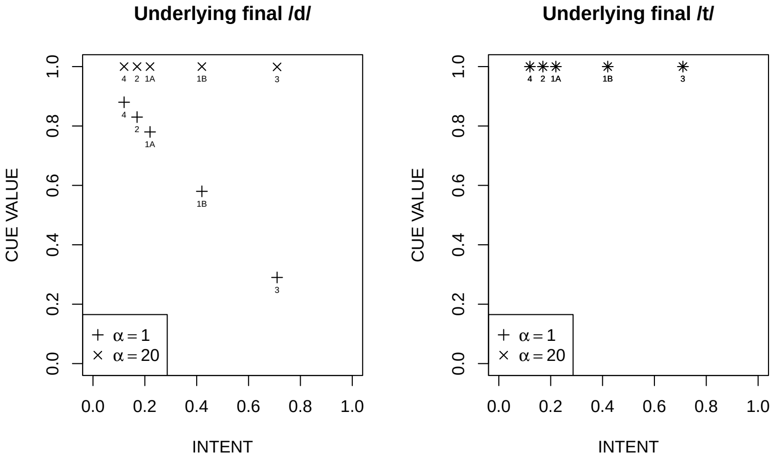

Figure 3: Simulated cue values for Port and Crawford (1989) results. Left plot shows values for /d/-final words and right plot shows values for /t/-final words. + represents burst duration and × represents proceeding vowel duration.

final /t/ segments. The vowel duration cue values for /d/ are close to 1 as they are for /t/. For /t/, burst duration is close to 1 as well. Again, this is expected given the data. While it may seem trivial that all of the underlying final /t/ values are right at 1 given that they were the denominator for determining ratios, these values were derived with the same formula that derived the final /d/ values. That is, the same α and same intent values were used, but the formula ensures that final /t/ values are unaltered.

The overall structure of the BLUEPRINT MODEL OF PRODUCTION 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)”. While this interpretation follows from conceptualizing the phonetics-phonology interface in terms of the modular feed-forward model, it does not follow from conceptualizing it in terms of the BLUEPRINT MODEL OF PRODUCTION. This is because the BLUEPRINT MODEL OF PRODUCTION 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 “rule”. Port and Crawford (1989) show that pragmatic context is a necessary ingredient, which is formalized in the BLUEPRINT MODEL OF PRODUCTION as intent. 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 continuous “intent”. It follows that under the BLUEPRINT MODEL OF PRODUCTION a continuous phonetic output does not require a continuous phonological grammar.

4.4 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, while 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 BLUEPRINT MODEL OF PRODUCTION.

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 we might expect 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: pɔnɡ22(55) → pɔnɡ35 'a scale'.[8] 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 /tɐk̚/ both surface with a mid-rising tone in contracted syllables:

(12) a. paŋ22 tsɔ35 → pɔ35 'to weigh (PERF)'
b. pɔŋ22 tɐk̚55 → pɔ35 'to weigh (POTENTIAL)

What makes it interesting is that /tsɔ/ has an underlying mid-rising tone while /tɐk̚/ 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 BLUEPRINT MODEL OF PRODUCTION, 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. We take the same approach as above 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 α 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 (/35/->[55]). Once again, the faithful mappings are unaffected by the α and intent values, and the values for the alternation mapping is an interpolation between these two extremes.

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

[8] Parenthesis represent a floating tone.

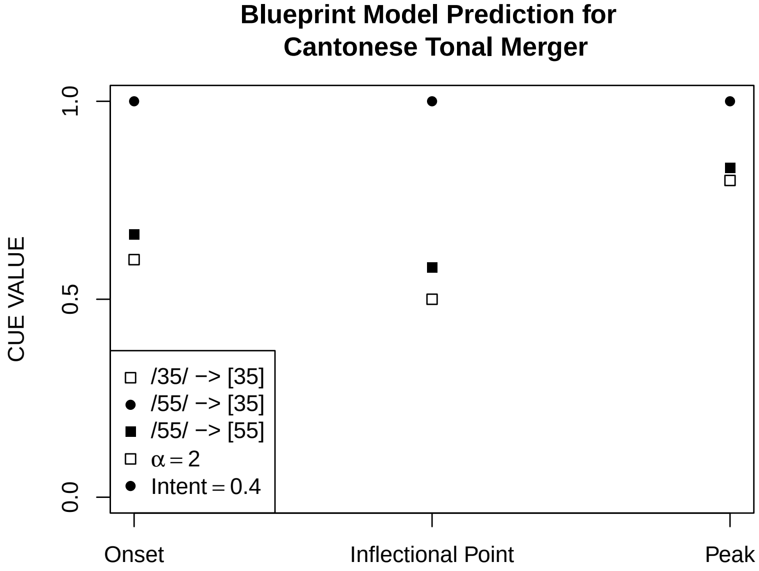

Figure 4: Simulated cue values for Yu (2007) tone-sandhi data

as two separate phenomena, they can, in certain cases, emerge from the same basic system. The BLUEPRINT MODEL OF PRODUCTION only relies on a correspondence between underlying and surface forms which is anticipated through the phonological mapping. Any phonological alternation, whether it be morphologically driven or otherwise, will predict the same type of phonetic effects in this model. The phonetic distribution of any segment should therefore be bounded between what we would expect given the underlying form and the actual surface form.

4.5 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 BLUEPRINT MODEL OF

PRODUCTION.

Since the BLUEPRINT MODEL OF PRODUCTION relies on a segmental correspondence between underlying and surface forms, the correspondant of an epenthesized segment is arguably zero. The spacial 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 average duration for an [i] vowel when *intent* was 0. If the level of *intent* is in between 0 and 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.[9] 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 three categories: 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 assigned the vector (450,2000) to indicate a more central vowel. Since there was more movement along the F2 dimension in the Hall (2013) data, the F2 cue was determined with an α value of 2 while F1 was determined with an α value of 2.4 (since a higher α 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 model the different type of speakers. For example, it is possible to have a single intent value and instead have the α levels for different cues vary across speakers. There is not enough empirical data to choose between the two modeling strategies here. Therefore, we again emphasize that this simulation is only one way to instantiate the BLUEPRINT MODEL OF PRODUCTION.

Another dimension that can affect the modeling results is the spatial parameters of the underlying zero form. This also varies drastically based on what choices are made in regards to phonetic

[9] Individual differences were not reported.

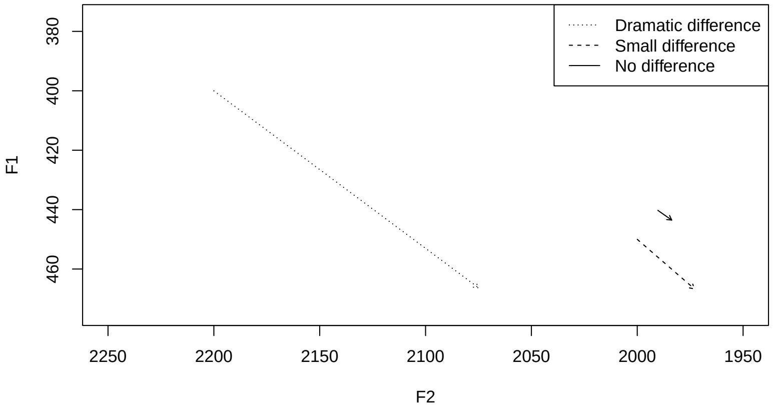

Figure 5: Simulated formant values for lexical and epenthetic [i] vowels based on Hall (2013) for a dramatic difference speaker, a small difference speaker, and a no difference speaker

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.

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 what the BLUEPRINT MODEL OF PRODUCTION predicts by having access to the lexicon, the phonological grammar, and the pragmatic context in which utterances are being made. The BLUEPRINT MODEL OF PRODUCTION 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). 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 BLUEPRINT MODEL OF PRODUCTION 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. 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: $A_{\text{BP}} :: L \to P \to I \to F \to \{PR\}$. All this tells us is that 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 can also be distinguished by syntactic and semantic information.

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 spacial and temporal information. Given π, 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 τ 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, δ maps frequencies to the unit interval. These functions π, τ and δ can be considered sub-programs within the larger phonetics function A_{BP}.

In order to have a full computational level description of the production process, we also need to describe how the various input elements interact. We propose that the target value output by the τ function is multiplied by the output of the δ function to provide a frequency scaled phonetic output. Following the assumption that the phonetic representation is a vector of parameters, the δ function outputs a vector rather than a scalar. In order to only influence the temporal cues, the output vector contains 0 at all of the indices corresponding to spacial cues. All other indices contain the value determined by the frequency function δ. Element wise multiplication would result in a phonetic representation where all the temporal cues have been scaled by the frequency determined factor. This shows that frequency effects can occur under the architecture of the BLUEPRINT MODEL OF PRODUCTION 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) model 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, ε is some random error, and λ 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 λ is the output of $\delta(F(L))$, a random error term can easily be added to any model, and instead of adding the bias term to the target, they are multiplied.

This implementation raises the question as to whether or not frequency information can influence non-temporal cues. For example, it is logically possible that the level of nasalization on vowels could also be affected by lexical frequency. We are not aware of anything like this. Regardless, the existence of such a phenomenon ultimately has little effect on the architecture of the BLUEPRINT MODEL OF PRODUCTION. Instead it would provide evidence for or against *particular instantiations* of the phonetics-phonology interface that fit the architecture of the BLUEPRINT MODEL OF PRODUCTION.

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.[10]

This simulation uses a linear model to predict the effect of frequency on duration. The model's outcome variable (y) is duration (ms) and has two predictor variables: log frequency and phonological form. This model structure 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 full linear model.

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

These parameters can be broken down to show how they relate to the functions above. Under the operating assumption that duration scales linearly with frequency, the underlying target

[10] 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.

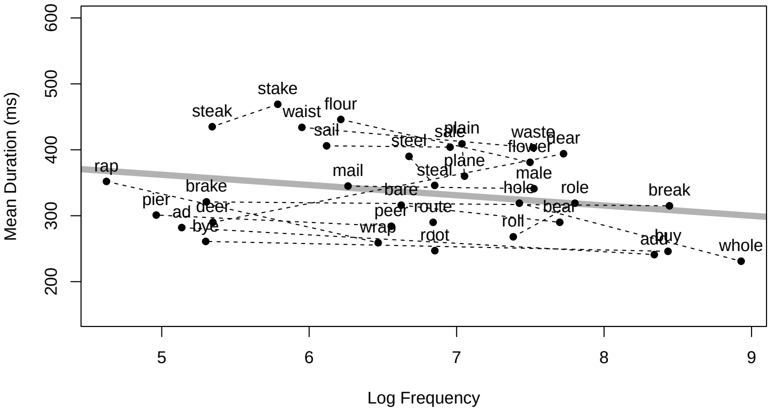

Figure 6: Average duration and log frequency for 17 words and their homophone twin. These data come from the Switchboard corpus (Godfrey et al., 1992). Dashed thin lines connect all homophonous pairs. The thick gray line is the output of a linear model of these points showing a general negative correlation.

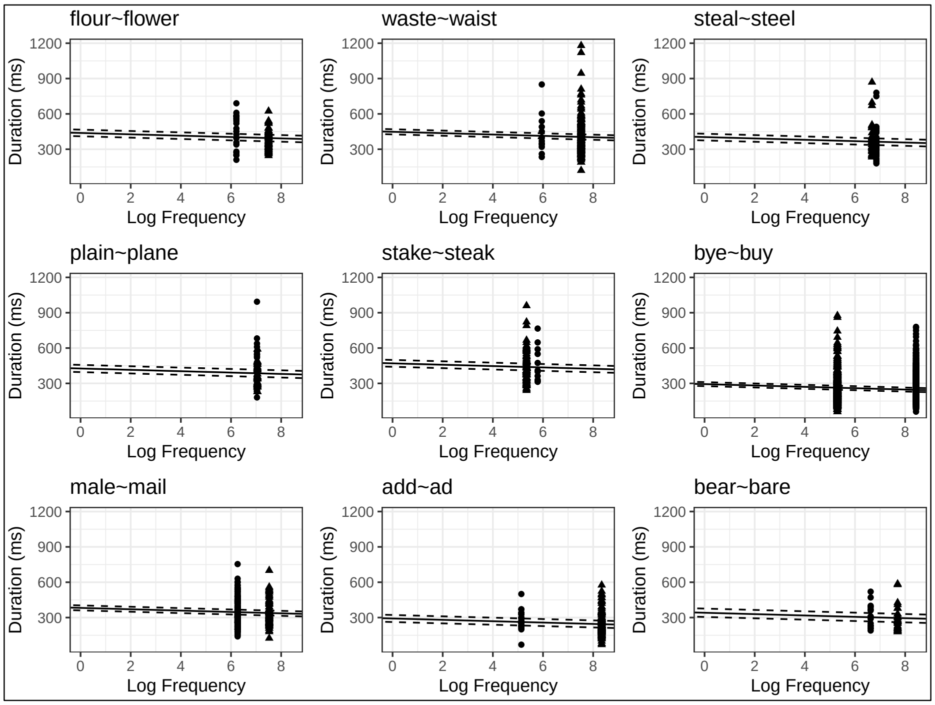

Figure 7: Frequency and Duration information for individual tokens of 9 randomly selected homophonous pairs. Each plot represents a single pair. The solid black lines are the predicted linear relationship for that phonological form. Dashed lines indicate 95% confidence intervals.

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.

To relate the linear model to the duration scaling function $\delta(F(L))$ above, it is necessary to do some rearranging of terms. In its current form, the linear model is more in line with the Pierrehumbert (2002) instantiation. 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 $y = \beta_1 \times LogFreq(x) + k$. Using basic algebra, an equivalent form is $y = k \times (1 + \frac{\beta_1 \times LogFreq(x)}{k})$. Since β_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 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 the linear model for each phonological form. The linear model has a significant effect for frequency (β = -5.761, t=-3.561, p < 0.001). To illustrate how this works, consider the pair bye∼buy. The model 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 effects 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 an error term in the model can be directly thought of as the aspects of production other than frequency that influence duration for a given production. Therefore, specific results of the linear model presented here should be interpreted conservatively.

Rather than focus on perfect prediction, the goal here was to show how the architecture of the

BLUEPRINT MODEL OF PRODUCTION can be used to model 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.

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 saturated one at at a time. 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 $A'^{l,p}_{BP} :: E \to \{PR\}$. 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.

Furthermore, this view highlights the importance of certain information over others during the production process. While all the factors play 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.[11] On the other hand, while pragmatic intent and lexical frequency influence the phonetic output, they do so by scaling the targets that are determined by the lexicon and phonology.

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

The role of phonetics in the Blueprint Model of Production 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

[11] 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.

provide the final set of instructions to the motor system. In this sense, the BLUEPRINT MODEL OF PRODUCTION 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 this paper, we show how the BLUEPRINT MODEL OF PRODUCTION is able to simulate 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 BLUEPRINT MODEL OF PRODUCTION. 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.

Braunschweiler, N. (1997). Integrated cues of voicing and vowel length in german: A production study. *Language and Speech*, 40(4):353–376.

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. (1999). Usage-based phonology. *Functionalism and formalism in linguistics*, 1:211–242.

Bybee, J. (2001). *Phonology and language use*, volume 94. Cambridge University Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3):129–159.

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.

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*.

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.

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.

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

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*.

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

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.

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

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., 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.

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.

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

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.

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.

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.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.

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.

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.

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.

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.

Vago, R. M. (1976). Theoretical implications of hungarian vowel harmony. *Linguistic Inquiry*, 7(2):243–263.

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

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.

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.