

Incomplete Neutralization and the Blueprint Model of Production

Scott Nelson and Jeffrey Heinz
Stony Brook University

1 Introduction

The discovery that phonologically neutralized segments vary in their phonetic form has been a problem for generative theories of phonology. Final devoicing is probably the most well studied example of this (Port et al., 1981; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985; Warner et al., 2004; Dmitrieva et al., 2010:*inter alia*). Phonologically, 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. It has been attested in a variety of languages including, but not limited to, German (Bloomfield, 1933), Polish (Gussmann, 2007), Catalan (Wheeler, 2005), Russian (Coats & Harshenin, 1971), and Turkish (Kopkalli, 1993). The data in (1) provide an example from German (Dinnsen & Garcia-Zamor, 1971).

- | | | | |
|-----|----------------------------------|-----|---|
| (1) | a. /bad+en/ → [baden] 'to bathe' | (2) | a. [-sonorant] → [-voice] / $_\sigma$ |
| | b. /bad/ → [bat] 'bath' | | b. IDENT-ONSET _{VOICE} >> *VDOBS >> IDENT _{VOICE} |
| | c. bat+en/ → [baten] 'asked' | | |
| | d. /bat/ → [bat] 'ask' | | |

This pattern has been straightforwardly accounted for with both rules (Grijzenhout, 2000; Wiese, 2000) and constraints (Lombardi, 1999) (see (2)). This is unsurprising as grammars based on rewrite rules such as those found in *The Sound Pattern of English* (SPE; Chomsky & Halle, 1968) and constraint-based grammars such as Optimality Theory (OT; Prince & Smolensky, 1993) can be thought of as two different ways to describe the same set of phonological mappings (Heinz, 2018). Both intensional descriptions described in (2) pick out the same set: voiceless obstruents that map faithfully to themselves regardless of syllabic position, and voiced obstruents that map faithfully when in onset, but map to their voiced counterparts when in coda position. Therefore, both rule and constraint grammars may be thought of as being *extensionally* equivalent when it comes to these data. This highlights the fact that the problem of incomplete neutralization may have nothing to do with the type of grammar doing the mapping and we should instead look elsewhere for a potential solution. If the phonological output is always a voiceless segment in coda position, then we may expect the phonetic output to be identical regardless of underlying form. Since the phonetic facts about final devoicing indicate that the outputs are not identical, models of language production that make no distinction between derived and underlying voiceless segments during implementation are unequipped to explain this phenomenon.

Traditionally, phonological mappings have been described as mappings from discrete underlying representations to discrete surface forms. This type of symbolic computation has historically been useful for describing sound patterns, but has been challenged as our understanding of the phonetic properties of language has increased over time (Ohala, 1990, 1992; Port & Leary, 2005). The architecture for language production is therefore often modeled in a way such that the phonetic implementation module only has access to the output of phonology and is therefore blind to any underlying contrast. This is what Pierrehumbert (2002) calls a "modular feed-forward" model. It is modular because each module operates over a specific input without any reference to a previous form. In the case of language production, this means that the phonetics module will have a surface representation (SR) from the phonology module as its input, but will

* We thank Ellen Broselow, Karthik Durvasula, Marie Huffman, and Bill Idsardi for helpful discussions regarding this material.

not be aware of what lexical item (or underlying representation (UR)) specifically lead to that SR. The model is feed-forward simply because there are no loops backwards. A visual representation of this type of model for speech production can be seen in Figure 1.

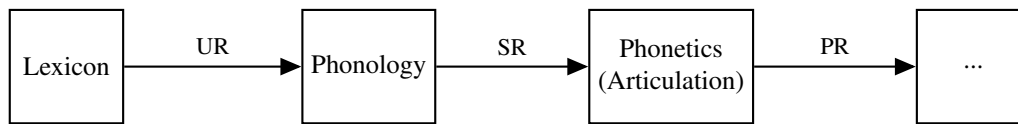


Figure 1: A modular feed-forward model of speech production. UR = underlying representation; SR = surface representation; PR = phonetic representation.

Given the phonetic facts about incomplete neutralization, there have been a variety of different modifications to this type of model in order to account for the data. Early proposals suggested that certain phonetic implementation rules happen before phonological rules (Dinnsen & Charles-Luce, 1984; Slowiaczek & Dinnsen, 1985). This requires potentially enriching phonological representations with phonetic information and therefore loosens both the feed-forward and modular aspect of the model. Van Oostendorp (2008) maintains feed-forward modularity and instead changes the representations used in the phonological grammar so that its output contains information about the lexical form. By doing so, the phonetics module can make a distinction between neutralized segments if need be. Other researchers reject this type of model either tacitly by using exemplar (Kleber et al., 2010; Winter & Röettger, 2011; Röettger et al., 2014) or dynamic (Gafos & Benus, 2006) theories of grammar, or explicitly in the case of Port & Leary (2005).

In this paper, we offer a new account of incomplete neutralization by reconceptualizing the modular feed-forward model into what we call the BLUEPRINT MODEL OF PRODUCTION. This is done at a computational level by conceptualizing the language production pipeline as a series of functions (Marr, 1982; Roark & Sproat, 2007; van Rooij & Baggio, 2021). By considering the types of each function (Pierce, 2002) it is shown that the modular feed-forward model can be rearranged in such a way that the phonetics module can have direct access to lexical information without making any modifications to the phonological module. Phonology remains a discrete function from underlying to surface forms that does not need to be enriched with phonetic information. The only change is that the phonological map becomes a direct input to the phonetics function, therefore conceptually acting as a set of directions (i.e. - blueprint) to the phonetics module for how a lexical item should be pronounced rather than something that directly creates the final pronounced form.

This paper is organized as follows. Section 2 discusses the facts of incomplete neutralization before introducing the BLUEPRINT MODEL OF PRODUCTION in detail in section 3. Section 4 compares the BLUEPRINT MODEL OF PRODUCTION to previous accounts of incomplete neutralization and discusses its implications for theories of the phonetics-phonology interface. Section 5 concludes the paper and briefly suggests ways in which this model can be used in future research.

2 Incomplete neutralization

In the 1980's, 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 segment varied over certain acoustic properties (Port et al., 1981; Port & O'Dell, 1985). Acoustically, it was found that the preceding vowel was shorter for underlying voiceless segments, the duration of aspiration noise was longer for underlying voiceless segments, and the amount of voicing into stop closure was longer for underlying voiced segments. These properties make it appear as if the surface form maintained some of the properties of the underlying form. Because the values for the derived voiceless segments were intermediate between a surface voiceless segment derived from underlying voiceless segment and a surface voiced segment (non-coda position), this phenomenon was termed “incomplete neutralization”.

Since these original landmark studies, final devoicing has been thoroughly studied, and found to be incomplete, in many other languages. In Catalan, there were no group level differences in the acoustics of derived and underlying voiceless segments, but due to individual level differences being found in the duration of voicing into closure, duration of closure, and length of previous vowel it was claimed that the process still remained incomplete (Dinnsen & Charles-Luce, 1984). Warner et al. (2004) found that, like in German, the

difference between segments that had been underlying [-voice] versus those that had been [+voice] in Dutch could be reliably measured in the acoustic signal as well as perceived by listeners. Preceding vowel length and closure duration were the most significant cues to the underlying voice contrast in both perception and production. Polish and Russian both showed acoustic signs of incomplete neutralization as well (Slowiaczek & Dinnsen, 1985; Dmitrieva et al., 2010; Shrager, 2012). For Polish it was the length of the preceding vowel that signified the underlying contrast, while in Russian it was burst duration. In Afrikaans, both closure duration and burst duration varied in the acoustics of the segments while burst duration and preceding vowel lengths aided in perceiving a difference between the underlying contrast (Van Rooy et al., 2003).

Incomplete neutralization is not limited to the process of final devoicing as the following processes have also been documented as being incomplete: tapping/flapping in American English (Fox & Terbeek, 1977; Herd et al., 2010; Braver, 2014), intrusive stops in English (Fourakis & Port, 1986), schwa epenthesis in English (Davidson, 2006), coda aspiration in Andalusian Spanish (Gerfen, 2002; Bishop, 2007), coda liquids in Puerto Rican Spanish (Simonet et al., 2008; Beaton, 2016), French schwa deletion (Fougeron & Steriade, 1997), Russian voicing assimilation (Burton & Robblee, 1997), epenthesis in Lebanese Arabic (Gouskova & Hall, 2009), tone sandhi in Cantonese (Yu, 2007), Mandarin (Peng, 2000) and Fuzhou (Li, 2016), Japanese monomoraic lengthening (Braver & Kawahara, 2016), and word-final lenition in Chilean Spanish (Bolyanatz, 2020). Durational cues appear to be the majority of what is presented as being incomplete, but there is not enough evidence to suggest it is limited to the temporal domain as in some cases it was exclusively non-durational cues which were found to be incomplete (Burton & Robblee, 1997; Yu, 2007).

Returning to final devoicing, (Port & Crawford, 1989) show 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 simply a wordlist.

They found incomplete neutralization in every condition when looking at combined speaker data. Interestingly, they found no difference between conditions 1A and 1B despite it previously being shown that repeating the target words back after hearing them spoken resulted in no signs of incomplete neutralization (Jassem & Richter, 1989). The most important takeaway from the results presented in Port & Crawford (1989) is the fact that 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", but it is not as necessary. Condition 3 meanwhile highlights the contrast, but provides no additional information. Now the speaker must use the amount of "voicing" if they want to make sure that 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 & 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.

To conclude this section, there are three major takeaways from the phonetic facts about incomplete neutralization. First, the direction of the incompleteness regularly appears to be in the direction of the underlying form. Second, it is often different cues that are incomplete. Third, speakers have at least some control over the level of incompleteness which is influenced by extra-linguistic factors. To account for these facts, a model of phonetic implementation must be able to reference both the underlying and surface form of a lexical item and have a way to integrate extra-linguistic information that can target specific cues rather than the entire phonetic output. In the next section, we introduce a model of the phonetics-phonology interface based on these insights.

3 The BLUEPRINT MODEL OF PRODUCTION

The biggest problem in accounting for incomplete neutralization comes from the generative view of language production as being a modular feed-forward process (Pierrehumbert, 2002:Figure 1). This type of model is like an assembly line: the intended lexical item moves down the line and gets modified along the way. To make this metaphor more concrete, the lexicon places URs on a conveyor belt which takes them to the Phonology station to be worked on. At the Phonology station, URs are transformed into SRs and SRs are placed back on the conveyor belt to be taken down the line to the Phonetics station. Noteworthy here is the fact that the Phonetics station will be receiving an SR with no knowledge of its previous history. Given a specific SR, the phonetics will transform it into the corresponding phonetic form (e.g. - a gradient representation containing acoustic/articulatory instructions). The data on incomplete neutralization presented in the previous section indicate that this view cannot be correct. That is, given the fact that there are cases where two segments are thought to be phonologically identical ($SR_i = SR_j$), but nonetheless have different phonetics (phonetic form_i \neq phonetic form_j), then the phonetics module (i.e. - “station”) must be distinguishing them based on some other piece of information.

The solution proposed here is to allow the phonetics module to have direct access to the underlying form. The remainder of this section outlines what we call the BLUEPRINT MODEL OF PRODUCTION which is a reimagining of the generative modular feed-forward model. It is modular in the sense that there are separate modules (minimally a phonology module and a phonetics module plus the lexicon). It is feed-forward in the sense that there are no feedback loops. Crucial to this analysis is the viewpoint that each module can be thought of as a function (Roark & Sproat, 2007; Heinz, 2018). In the modular feed-forward model we might think of the *Phonology* as a function that maps a *UR* to an *SR* and the *Phonetics* as a function that maps an *SR* to a *PR* (Phonetic Representation). The BLUEPRINT MODEL OF PRODUCTION continues to view the *Phonology* as a function that maps a *UR* to an *SR* but models the *Phonetics* as a higher order function that takes the *Phonology* function as an input. Instead of a single *UR*, the entire *Lexicon* is also an input to the *Phonetics* function in order to generalize over all lexical items. The final input is a scaling value called *Intent* which represents certain extra-linguistic information. The *Phonetics* module is therefore a function with three inputs: *Lexicon*, *Phonology*, and *Intent*; and one output: $\{PR\}$. A visual representation of the model is shown in Figure 2.

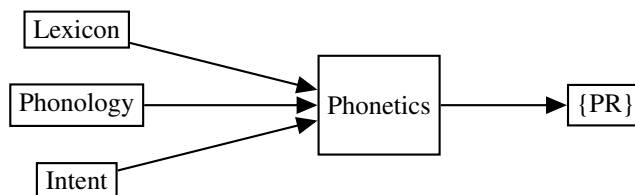


Figure 2: A visual representation of the BLUEPRINT MODEL OF PRODUCTION

3.1 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 modular feed-forward model, it is actually a straightforward reconceptualization when viewed as a series of mathematical functions. The analysis below shows the types of the functions following Pierce (2002) whose notation is derived from the lambda calculus (Church, 1932, 1933). The general form that will be used to describe functions is $f :: A \rightarrow B$ which can be read as f is a function that maps A -type arguments to B -type arguments. The *Phonology* function (or P for short) above would therefore be written as $P :: UR \rightarrow SR$. In prose this would be equal to, “the phonology function P maps URs to SRs.”

The analysis below also relies on two other concepts: higher order functions and the notion of function application. A higher order function is a function that either takes another function in its input, or a function that returns another function in its output. One function of the first type is the *map()* function.¹ Given two inputs A and B , where A is a function and B is an array of length n , $map(A, B)$ will apply function A to every individual element of B and return the array $[A(B_0), \dots, A(B_n)]$. To give a concrete example, consider

¹ This function can be found natively in programming languages such as Python, Haskell, Ruby, Perl, and R.

the function *plus-one()* 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 $\text{map}(\text{plus-one}(), [-23, 1, 9, 307]) = [-22, 2, 10, 308]$. 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 *b-if-c()* that turned any *c* into a *b*, then $\text{map}(\text{b-if-c}(), abc) = abb$. The larger point here is that a function's status of being higher-order or not is a separate issue from exactly what the types it operates over are.

This point leads us nicely into a discussion of function application which can itself be thought of as a higher order function. Abstracting away from types, the two arguments for *function-application()* itself would be one of type *A* and the other of type $A \rightarrow B$ (i.e. - a function that maps *A* type things to *B* type things). Given these two arguments it would output something of type *B*. The type for function application would therefore be $A \rightarrow (A \rightarrow B) \rightarrow B$. Note that when reading types, 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. In a system where we know all *B* type things are derived from *A* type things, and that there is some function *g* that implements this mapping, then we also know that every *B* is equivalent to $g(a)$ for some $a \in A$.

Let us return to the model of language production. Throughout the remainder of this section I will use the following abbreviations: *L*, *P*, and *A* as functions representing the **Lexicon**, **Phonology**, and **Phonetics** (Articulation since we are modeling production); *UR*, *SR*, and *PR* to represent **Underlying Representations**, **Surface Representations**, and **Phonetic Representations**; and *I* to represent **Intent** on maintaining an underlying contrast. The proposed types are listed in the table in Figure 3a. Figure 3b explains how the modular feed-forward model leads to the BLUEPRINT MODEL OF PRODUCTION.

Name	Meaning	Type	
<i>L</i>	Lexicon	$UR \rightarrow UR$	
<i>P</i>	Phonology	$UR \rightarrow SR$	1. $A_{\text{MFF}} :: SR \rightarrow PR$
A_{MFF}	Phonetics _{MFF}	$SR \rightarrow PR$	2. $A :: UR \rightarrow (UR \rightarrow SR) \rightarrow PR$
A_{BP}	Phonetics _{BP}	$L \rightarrow P \rightarrow I \rightarrow PR$	3. $A :: UR \rightarrow P \rightarrow PR$
<i>UR</i>	Underlying Rep.	String (Feature Vectors)	4. $A :: L \rightarrow P \rightarrow \{PR\}$
<i>SR</i>	Surface Rep.	String (Feature Vectors)	5. $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$
<i>PR</i>	Phonetic Rep.	Matrix (\mathbb{R})	
<i>I</i>	Intent	$\mathbb{R} \in [0, 1]$	

(a) Types

(b) The derivation

Figure 3: Deriving the Blueprint Model

Let us start by analyzing the phonetics module in the modular feed-forward model (Step 1 in Figure 3b). 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 $UR \rightarrow (UR \rightarrow SR)$ (Step 2). We can now further switch out $(UR \rightarrow SR)$ since we know that this is just another way of representing the phonology module (Step 3). To complete this reconceptualization we can do two more things. First, we can change *UR* to *L* in order to generalize over the entire lexicon (Step 4) (Roark & Sproat, 2007). By doing so, we also make the output a set of phonetic representations rather than a single specific representation. Second, we can add in *I* as an input in order to account for pragmatic effects. This gives us our final type for the phonetics function shown in (Step 5).

The phonetics module is therefore a higher-order function with three arguments: the lexicon, the entire phonology module (a function), and an intent value. As is the case in the modular feed-forward model, the phonology 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 where the phonology processes the underlying form into a surface form. In the modular feed-forward model it is external to the phonetics module while in the BLUEPRINT MODEL OF PRODUCTION it is internal 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 the module that

actually creates a phonetic output, but instead as a set of instructions that informs the phonetics module as to how a given input should be pronounced. In other words, the phonetics module queries the phonology as to how a UR should be pronounced by determining the correct $\langle UR, SR \rangle$ pair. Once it has this information, other factors (such as contrastive intent) may further influence the final phonetic representation. In the case of incomplete neutralization, it appears that the phonetic representation is a blend of the phonological output (SR) and the lexical input (UR) that is scaled by extra linguistic factors (I).

3.2 Scaling UR/SR and pragmatics Recall that the direction and scalability of incomplete neutralization have influenced the development of the BLUEPRINT MODEL OF PRODUCTION. 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 & Crawford (1989) showed that participants seemed to scale the level of incompleteness based on pragmatic context. Taken together, these two empirical findings suggest that incomplete neutralization is the result of scaling the phonetic representation in order to be more like the underlying form.

Imagine a one-dimensional space for some cue c in the set of all cues C that signify a contrast like voicing. We can divide the space up 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 we may even get 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 must now also be the case that the [-voice] sub-section can also 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.² Therefore, devoiced obstruents in final position can be mapped to all values between m and n in the cue space as a result of the scaling described above.

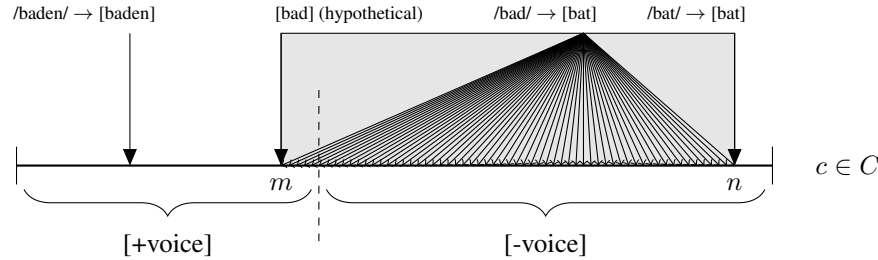


Figure 4: Cue space

From the basic idea sketched above, we can now formalize this process. First, the scaling value can be formalized as *intent* following Gafos & Benus (2006). This can be thought of as the percentage that a speaker wants to maintain the underlying contrastive form. Numerically, we can think of this as a value $i \in [0, 1]$. The lower bound 0 represents a speaker with 0% intent to maintain the underlying contrast while the upper bound 1 represents a speaker who wants to 100% maintain the underlying contrast. We can then figure out exactly what the value for cue c_i by simply taking the weighted sum of c_{UR} and c_{SR} . In the example for devoicing $c_{UR} = m$ and $c_{SR} = n$. An exact formula for calculation c_i is given in (3).

$$(3) \quad c_i = c_{UR} \times i + c_{SR} \times (1 - i)$$

² 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. Anecdotaly, speakers of languages with final devoicing can produce a word final obstruent as voiced if absolutely forced to do so. Additionally, based on our own dialects of American English, which neutralizes /t/ and /d/ to [ɾ] in certain positions, we can produce words like *waiter* with a [t] or at the very least [ɾ].

The value for a cue will therefore always be what the c_{SR} value is unless the speaker has some desire to maintain the underlying contrast. Notice, too, that there will only be a change in the expected cue value when there is some alternation between underlying and surface form. Since the intent value is some real number between $[0, 1]$ then it will always be the case that $i + (1 - i) = 1$. This means that whenever $c_{UR} = c_{SR}$ there is no way to not end up with $c_i = c_{SR}$, which is exactly the desired result given we only want *intent* to play a role when there is an alternation.

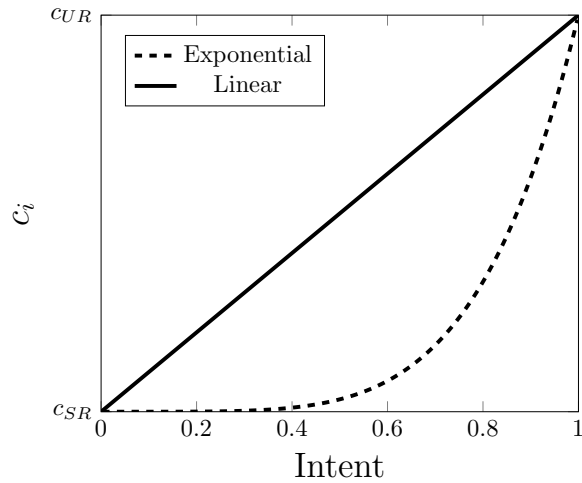
The current formulation in (3) suggests that the weighting of UR and SR is scaled linearly. It is quite likely that this is not the correct formulation since all of the documented phonetic effects of incomplete neutralization are subtle. Given the current formulation, we would expect to see more intermediate cue values. Scaling the UR and SR values exponentially rather than linearly helps solve this problem. This would mean that for smaller intent values there would be little effect on the cue output, but for larger values it would be more noticeable. Since there is anecdotal evidence that speakers can produce the UR value if they actively try, this seems to indicate that it is only with extremely high values of intent that the cue output should be able to cross the threshold into an area that is reliably interpreted as the UR segment. The term α further scales the intent values, giving us (4) below.³

$$(4) \quad c_i = c_{UR} \times i^\alpha + c_{SR} \times (1 - i)^\alpha$$

The difference between linear weighting of intent values and exponential weighting of intent values is shown in Figure 5b. Here, it is apparent that c_i will only pass the halfway point between c_{UR} and c_{SR} for intent values close to 1.

The following example illustrates the scaling process. In Warner et al. (2004), the difference between preceding vowel duration in the environment where voicing in stops is distinct was around 20ms. In the neutralizing environment, underlying voiceless stops had a preceding vowel duration of 120ms while underlying voiced stops that were neutralized to voiceless had a preceding vowel duration of 124ms. Given that the preceding vowel before voiced stops is typically longer, we can hypothesize that the preceding vowel duration target for a word final [+voice] stop in Dutch is 140ms (20ms > than the [-voice] stop in the same position). The table in Figure 5a shows the predicted preceding vowel length values for a neutralized underlying voiced stop with intent values incrementing by 0.1 from 0 to 1 using the formula in (4).

Intent	Vowel Duration (ms)
0	120
0.1	120
0.2	120.01
0.3	120.05
0.4	120.2
0.5	120.62
0.6	121.56
0.7	123.36
0.8	126.55
0.9	131.81
1	140



(a) Predicted preceding vowel durations

(b) Linear vs. Exponential weighting of intent value

Figure 5: Exponentially scaling vowel duration (left) and Linear vs. Exponential weighting (right)

With intent values less than 0.6, the difference in vowel length before a derived [-voice] segment and an underlying [-voice] segment is predicted to be less than 0.7 milliseconds. This small of a difference

³ For most of this paper $\alpha = 5$ is arbitrarily used. It should be noted that as the value of α in i^α increases, the output stays closer to 0 longer. Determining the exact value of α has no bearing on the larger point being made, but may be an interesting way to explain variation in what types of cues and processes exhibit “incomplete” behavior.

would certainly go unnoticed both in casual conversation and even in laboratory settings. Furthermore, even with moderate intent values (0.6-0.7) the difference still remains quite small, but enough to result in statistically significant differences. In fact, the predicted preceding vowel duration with *intent* value equal to 0.7 is 123.36ms which is close to the measured average of 124ms reported by Warner et al. (2004). Once the intent value goes between 0.8-1 the predicted vowel duration value drastically becomes more like the UR value ([+voice]). The equation in (4) is therefore able to match empirical findings. Furthermore, the measurements in Warner et al. (2004) were taken from word *list* recordings. Port & Crawford (1989) reported that word list read speech showed a moderate level of incompleteness in German. In the simulation shown in Table 5a, there is still room above the 0.7 *intent* value for a task like the contrastive condition in Port & Crawford (1989) that showed a higher level of incompleteness.

4 Discussion

Up to this point the BLUEPRINT MODEL OF PRODUCTION has been primarily displayed on its own terms, but it is also necessary to compare it to other production models and how they account for incomplete neutralization. A goal in developing this model was to see if there was a way to maintain the following three conditions: a discrete phonology, modularity, and feed-forwardness. To start, let us examine an early proposal where certain phonetic implementation rules take place before certain phonological rules (Dinnsen & Charles-Luce, 1984; Slowiaczek & Dinnsen, 1985). One way to interpret this type of proposal is through the elimination of the feed-forward requirement. The problem with this interpretation is that classic definitions of modularity assume that different modules necessarily have unique representations (Fodor, 1983). If the phonetic implementation rules make direct reference to duration of stop closure (in ms) of a segment (e.g. Slowiaczek & Dinnsen (1985)), and then a phonological rule is implemented afterwards, then the phonology must be able to interpret and process the segment in its rich phonetic representation. This means that having phonetic implementation and phonological rules intertwine further leads to the degradation of modularity and potentially to discrete representations as well. The BLUEPRINT MODEL OF PRODUCTION is still a feed-forward system and therefore continues to maintain a difference between phonological and phonetic representations while still accounting for the facts of incomplete neutralization.

Another production model is that of turbidity theory (Goldrick, 2000) which extends the idea of containment as originally proposed in classic Optimality Theory Prince & Smolensky (1993:OT). In this version of OT, each segment has a “project” and “pronounce” relation. Because of this, information about the input can be contained within the output (projection) even if it is not pronounced. Van Oostendorp (2008) analyzes final devoicing within this framework. He proposes that the final voiced obstruent is in a projection relationship with the [voice] feature, but not in a pronounce relationship. The project relationship thus alters the voicing implementation slightly in the phonetics module. As Braver (2019) points out, this account is unable to explain Japanese monomoraic lengthening or American English flapping.

It is worth mentioning that the failure to account for these processes does not have to do with generally referencing information about the UR, but rather specifically how turbidity theory references UR information. Incomplete neutralization is the result of both a project and pronounce relation in the UR turning into solely a project relation in the SR. This can be thought of as another way of formulating delinking in an autosegmental framework. Japanese monomoraic lengthening on the other hand can be thought of as spreading. In this instance, a pronunciation relation is added between the segmental properties of one mora to the next. Similarly, the voicing of the flap would be addition of a pronunciation relation. This is a fundamentally different than what happens in Van Oostendorp’s (2008) account of final devoicing. The BLUEPRINT MODEL OF PRODUCTION on the other hand only cares about the mapping from UR to SR. It does not matter if it is a “de-linking” process or a “spreading” process, as long as there is a change from UR to SR then there is the possibility for incomplete neutralization.

Van Oostendorp’s (2008) account also raises a question that should be asked of any phonologist positing a theory that accounts for within-category variation: what is the role of a phonological grammar in production? Most accounts working off of the modular feed-forward model seem to assume that the phonological grammar is implementing the changes it makes to lexical items. With this type of interpretation, it does make some sense to add systematic within-category variation to the phonological grammar. But another valid interpretation is that the phonological grammar is just one of several factors impacting production. The phonology, in this type of interpretation, still operates as a symbolic manipulator of lexical items, and tells

the production module how to do so. It is this interpretation of the phonological grammar that is present in the BLUEPRINT MODEL OF PRODUCTION.

So, even though Van Oostendorp's (2008) account of incomplete neutralization using turbidity theory does integrate information about underlying and surface forms, it is still quite philosophically and structurally different from the BLUEPRINT MODEL OF PRODUCTION's account. It is a strength of the BLUEPRINT MODEL OF PRODUCTION that it is relatively agnostic about the form of the phonology function while still providing insight into incomplete neutralization.

Another model of production that accounts for incomplete neutralization and also falls under the OT umbrella is found in Braver (2019). He combines output-output correspondence (Benua, 1997; Steriade, 2000), phonetically motivated constraints (Zsiga, 2000; Flemming, 2001), and weighted constraints (Legendre et al., 1990) into what he calls WEIGHTED PARADIGM UNIFORMITY (WUP). For this account, the phonetic output is determined by a markedness constraint that requires the output to be identical to the canonical output and a faithfulness constraint that requires words within the same morphological paradigm to have identical outputs. The final phonetic output is therefore a blend of the canonical output (the phonological SR) and the base of the morphological paradigm. As he outlines (pp. 14–19), there are many different criteria that have been used to determine bases. He ultimately uses informativeness in his analysis of Japanese monomoraic lengthening, but also discusses morphological complexity, frequency, and orthography as other factors influencing base decision. The WUP is similar to the BLUEPRINT MODEL OF PRODUCTION in that they both provide upper and lower bounds for the final phonetic output and can in theory account for gradient amounts of incompleteness. What makes them immediately different is that Braver (2019) relies on morphological bases while the BLUEPRINT MODEL OF PRODUCTION uses underlying forms, which are the basis of generative models of phonology.

Articulatory Phonology (AP; Browman & Goldstein, 1986, 1992, 1995) is a theory of production which collapses phonetics and phonology into a single source, unlike other theories (Kingston & Diehl, 1994; Tucker & Warner, 2010). AP claims 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. 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 & 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 what we consider phonological and phonetic processes are the lawful consequence of interacting gestures within a dynamical system.

Gafos (2002) integrated the gestural and dynamical approach into an OT grammar in order to account for non-allophonic phonological knowledge. This move maintains a distinction between phonetics and phonology in a way that classical AP did not. Gafos & Benus (2006) use this type of model to explain incomplete neutralization in final devoicing. The OT grammar corresponds to a dynamical equation and makes it so there is a canonical "voiceless" dynamical equation for glottal aperture (the gesture that corresponds to voicing in the task dynamic model). As discussed previously, they add an *intent* value that is another term in the dynamical equation. This has the property that as *intent* increases, the dynamical equation is adjusted so that the phonetic output becomes more "voiced" like.

The influence of Gafos & Benus (2006) on the BLUEPRINT MODEL OF PRODUCTION should be quite obvious. If incomplete neutralization is scalable based on pragmatics then the *intent* value seems necessary for any production model. Any account of incomplete neutralization has to either take this fact seriously or reject the results of Port & Crawford (1989). The BLUEPRINT MODEL OF PRODUCTION shows that this is possible to account for this fact without nonlinear dynamics. In general, a specific theory of phonology is not required to account for incomplete neutralization and instead the important point can be made at a more abstract level. Gafos & Benus (2006) appear to recognize this fact when they 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). This suggests the primary difference between the analysis in this paper and Gafos & Benus's (2006) analysis is that the BLUEPRINT MODEL OF PRODUCTION provides an abstract account of incomplete neutralization using functions and lambda calculus at a computational level while they provide a mathematical account of

incomplete neutralization using nonlinear dynamics at an algorithmic level (cf. Marr, 1982). Both types of analyses are necessary for understanding the phenomenon, but the BLUEPRINT MODEL OF PRODUCTION may ultimately appeal to a wider group of phonologists since it does not require an acceptance of gestural approaches or nonlinear dynamics.

Exemplar models have also been popular in accounting for incomplete neutralization (Kleber et al., 2010; Winter & Röttger, 2011; Röttger et al., 2014) (but see (Seyfarth et al., 2019)). In these types of models, the lexicon contains both roots and inflected forms for all lexical items. Additionally, the representations are phonetically rich and contain traces of all previously heard tokens of each item. The organization of the lexicon in this way can account for incomplete neutralization through spreading activation. What this means is that when a root form such as /ʁad/ in German is activated, so are all of the morphologically related forms (such as /ʁad+en/. The output of the final /d/ is a combination of how it is pronounced in the root and all the related morphological forms. In the case of final devoicing, the /d/ in /ʁad+en/ surfaces as a [d] and therefore influences the /d/ in /ʁad/ to be more [d]-like, even though the exemplar representation for /ʁad/ ends in a [t].

One strength of exemplar models is their ability to explain sub-phonemic variation. This, of course, is unsurprising given its detailed representations. The advantage of the BLUEPRINT MODEL OF PRODUCTION over exemplar based models when it comes to incomplete neutralization is the fact that it can do so without relying on detailed representations. Another difference is that exemplar models rely on morphological paradigms while the BLUEPRINT MODEL OF PRODUCTION only uses the underlying representation. This means that we should be able to find a case of neutralization where the two models make different predictions about the phonetic form, which we leave for future work.

As this section has shown, the BLUEPRINT MODEL OF PRODUCTION's advantage over other accounts of incomplete neutralization is its ability to maintain a distinction between phonetics and phonology. Unlike other accounts it does not require phonetically rich or gradient phonological representations. Phonological representations remain discrete and symbolic while phonetic representations are continuously valued. The BLUEPRINT MODEL OF PRODUCTION also does not require any feedback so it maintains the feed-forward aspect. That being said, it is not feed-forward in the same way that the modular feed-forward model is. There is no longer a temporal separation between phonetics and phonology despite them fundamentally providing different types of knowledge. Finally, the BLUEPRINT MODEL OF PRODUCTION is a mathematical reinterpretation of the modular feed-forward model rather than a complete abandonment. This means that any insights gained from using the more traditional model can also be used with this updated version. Accounts that reimagine the production process by merging phonetics and phonology (e.g. - exemplar, nonlinear dynamics, etc...) do not have this same luxury.

5 Conclusion

This paper offered a new model of language production that can account for incomplete neutralization without eliminating discrete phonological knowledge. While other accounts of incomplete neutralization have made similar claims, the BLUEPRINT MODEL OF PRODUCTION provides a mathematical treatment that shows how this type of model is simply a reinterpretation of the more classic modular feed-forward model rather than a complete reimagining of the production process. As currently presented, it explicitly recognizes three separate factors influencing production. The lexical representation is arguably the primary factor; the phonology is arguable a secondary factor; and the intent is arguably a tertiary factor. By providing an abstract model that does not rely on any specific theory of phonology it can hopefully be adapted by a wide range of researchers. The paper also provided more specific examples in how the new model could be implemented and used to simulate and explain different types of incomplete neutralization and near merger data. That being said, despite incomplete neutralization being the primary focus in the creation of the BLUEPRINT MODEL OF PRODUCTION, future work can use this model to explore other facts about the phonetics-phonology interface that relate to the phonetic realization of phonological processes involving segmental changes (e.g. epenthetic vowel duration, phonetic evidence of deleted segments, absolute neutralization).

References

- Beaton, Mary Elizabeth (2016). Revisiting incomplete neutralization: The case of Puerto Rican Spanish. *University of Pennsylvania Working Papers in Linguistics* 22:1, p. 5.
- Benua, Laura (1997). *Transderivational identity*. Ph.D. thesis, University of Massachusetts Amherst.

- Bishop, Jason B (2007). Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration. *Proceedings of the International Congress of Phonetic Sciences*, vol. 16, 1765–1768.
- Bloomfield, Leonard (1933). *Language*. University of Chicago Press.
- Bolyanatz, Mariška A (2020). Evidence for incomplete neutralization in Chilean Spanish. *Phonetica* 77:2, 107–130.
- Braver, Aaron (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua* 152, 24–44.
- Braver, Aaron (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology* 36:1, 1–36.
- Braver, Aaron & Shigeto Kawahara (2016). Incomplete neutralization in Japanese monomoraic lengthening. Albright, Adam & Michelle A. Fullwood (eds.), *Proceedings of the 2014 Meeting on Phonology*.
- Browman, Catherine P & Louis Goldstein (1992). Articulatory phonology: An overview. *Phonetica* 49:3–4, 155–180.
- Browman, Catherine P & Louis M Goldstein (1986). Towards an articulatory phonology. *Phonology* 3, 219–252.
- Browman, Catherine P & Louis M Goldstein (1995). Dynamics and articulatory phonology. Port, Robert F. & T. V. Gelder (eds.), *Mind as Motion*, MIT Press, 175–194.
- Burton, Martha W. & Karen E. Robblee (1997). A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25:2, 97–114.
- Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. Harper & Row.
- Church, Alonzo (1932). A set of postulates for the foundation of logic. *Annals of mathematics* 346–366.
- Church, Alonzo (1933). A set of postulates for the foundation of logic. *Annals of mathematics* 839–864.
- Coats, Herbert S & Alex P Harshenin (1971). On the phonological properties of Russian. *The Slavic and East European Journal* 15:4, 466–478.
- Davidson, Lisa (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34:1, 104–137.
- Dinnsen, Daniel & Maria Garcia-Zamor (1971). The three degrees of vowel length in German. *Research on Language & Social Interaction* 4:1, 111–126.
- Dinnsen, Daniel A. & Jan Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12:1, 49–60.
- Dmitrieva, Olga, Allard Jongman & Joan Sereno (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics* 38:3, 483–492.
- Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:1, 7–44.
- Fodor, Jerry A (1983). *The modularity of mind*. MIT press.
- Fougeron, Cécile & Donca Steriade (1997). Does deletion of french schwa lead to neutralization of lexical distinctions? *EUROSPEECH*, vol. 2, 943–946.
- Fourakis, Marios & Robert Port (1986). Stop epenthesis in English. *Journal of Phonetics* 14:2, 197–221.
- Fox, Robert A. & Dale Terbeek (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics* 5:1, 27–34.
- Gafos, Adamantios I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20:2, 269–337.
- Gafos, Adamantios I. & Stefan Benus (2006). Dynamics of phonological cognition. *Cognitive Science* 30:5, 905–943.
- Gerfen, Chip (2002). Andalusian codas. *Probus* 14:2, 247–277.
- Goldrick, Matthew (2000). Turbid output representations and the unity of opacity. *North East Linguistics Society* 30.
- Gouskova, Maria & Nancy Hall (2009). Acoustics of epenthetic vowels in Lebanese Arabic. *Phonological argumentation: Essays on evidence and motivation* 203–225.
- Grijzenhout, Janet (2000). Voicing and devoicing in English, German and Dutch: Evidence for domain specific identity constraints. *Theory des Lexikons; Arbeiten des Sonderforschungsbereichs* 282, 1–22.
- Gussmann, Edmund (2007). *The phonology of Polish*. Oxford University Press.
- Heinz, Jeffrey (2018). The computational nature of phonological generalizations. Hyman, Larry & Frans Plank (eds.), *Phonological Typology, Phonetics and Phonology*, De Gruyter Mouton, chap. 5, 126–195.
- Herd, Wendy, Allard Jongman & Joan Sereno (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics* 38:4, 504–516.
- Jassem, Wiktor & Lutoslaw Richter (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17:4, 317–325.
- Kingston, John & Randy L Diehl (1994). Phonetic knowledge. *Language* 70:3, 419–454.
- Kleber, Felicitas, Tina John & Jonathan Harrington (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38:2, 185–196.
- Kopkalli, Handan (1993). *A phonetic and phonological analysis of final devoicing in Turkish*. Ph.D. thesis, University of Michigan.
- Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic grammar—a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the twelfth annual conference of the Cognitive Science Society*, Citeseer, 884–891.

- Li, Yang (2016). Complete and incomplete neutralisation in Fuzhou tone sandhi. *Proc. 5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, 116–120.
- Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language & Linguistic Theory* 17:2, 267–302.
- Marr, David (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
- Ohala, John J (1990). There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18:2, 153–172.
- Ohala, John J (1992). The costs and benefits of phonological analysis. Downing, P., S. Lima & M. Noonan (eds.), *Literacy and Linguistics*, John Benjamins, 211–238.
- Peng, Shu-Hui (2000). Lexical versus ‘phonological’ representations of Mandarin sandhi tones. Broe, Michael B & Janet B. Pierrehumbert (eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, Cambridge University Press, 152–167.
- Pierce, Benjamin C (2002). *Types and programming languages*. MIT press.
- Pierrehumbert, Janet (2002). Word-specific phonetics. Gussenhoven, C. & N. Warner (eds.), *Laboratory phonology 7*, De Gruyter Mouton, 101–139.
- Port, Robert & Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17:4, 257–282.
- Port, Robert, Fares Mitleb & Michael O’Dell (1981). Neutralization of obstruent voicing in German is incomplete. *The Journal of the Acoustical Society of America* 70, p. S13.
- Port, Robert F & Adam P Leary (2005). Against formal phonology. *Language* 81:4, 927–964.
- Port, Robert F & Michael L O’Dell (1985). Neutralization of syllable-final voicing in German. *Journal of phonetics* 13:4, 455–471.
- Prince, A & Paul Smolensky (1993). Optimality theory: Constraint interaction in generative grammar (technical report ruccs-tr-2). Tech. rep., Rutgers University Center for Cognitive Science.
- Roark, Brian & Richard Sproat (2007). *Computational approaches to morphology and syntax*. Oxford University Press.
- Röetger, Timo B, Bodo Winter, Sven Grawunder, James Kirby & Martine Grice (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43, 11–25.
- van Rooij, Iris & Giosuè Baggio (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science* 16:4, 682–697.
- Saltzman, Elliot L. & Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1:4, 333–382.
- Seyfarth, Scott, Jozina Vander Klok & Marc Garellek (2019). Evidence against interactive effects on articulation in Javanese verb paradigms. *Psychonomic Bulletin & Review* 26:5, 1690–1696.
- Shrager, Miriam (2012). Neutralization of word-final voicing in Russian. *Journal of Slavic Linguistics* 20:1, 71–99.
- Simonet, Miquel, Marcos Rohena-Madrazo & Mercedes Paz (2008). Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. Colantoni, Laura & Jeffrey Steele (eds.), *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, Cascadia Proceedings Project, 72–86.
- Slowiaczek, Louisa M. & Daniel A. Dinnsen (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13:3, 325–341.
- Steriade, Donca (2000). Paradigm uniformity and the phonetics-phonology boundary. Broe, Michael B. & Janet B. Pierrehumbert (eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, Cambridge University Press, 313–334.
- Tucker, Benjamin V. & Natasha Warner (2010). What it means to be phonetic or phonological: the case of Romanian devoiced nasals. *Phonology* 27:2, 289–324.
- Van Oostendorp, Marc (2008). Incomplete devoicing in formal phonology. *Lingua* 118:9, 1362–1374.
- Van Rooy, Bertus, Daan Wissing & Dwayne D. Paschall (2003). Demystifying incomplete neutralisation during final devoicing. *Southern African Linguistics and Applied Language Studies* 21:1-2, 49–66.
- Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32:2, 251–276.
- Wheeler, Max W. (2005). *The phonology of Catalan*. Oxford University Press.
- Wiese, Richard (2000). *The phonology of German*. Oxford University Press.
- Winter, Bodo & Timo Röetger (2011). The nature of incomplete neutralization in German: Implications for laboratory phonology. *Grazer Linguistische Studien* 76, 55–74.
- Yu, Alan CL (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24:1, 187–214.
- Zsiga, Elizabeth C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28:1, 69–102.