Volume 121, Issue 8, June 2011

ISSN 0024-3841

ELSEVIER

Lingua

International Review of General Linguistics

Lingua available on ScienceDirect from Volume 1, published 1949

Available online at www.sciencedirect.com

ScienceDirect

Lingua 121 (2011) 1344–1358

Contents lists available at ScienceDirect

Lingua

journal homepage: www.elsevier.com/locate/lingua

Perceptual similarity in input–output mappings: A computational/experimental study of non-native speech production

Jason A. Shaw [a,b,*], Lisa Davidson [b]

[a] MARCS Auditory Laboratories/School of Humanities and Languages, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia
[b] New York University, 10 Washington Place, New York, NY 10003, United States

ARTICLE INFO

Article history:
Received 17 January 2011
Received in revised form 1 March 2011
Accepted 3 March 2011
Available online 13 April 2011

Keywords:
P-map hypothesis
Faithfulness
Phonotactics
Perceptual epenthesis
Cross-language speech perception/production
Stochastic modeling
Motor coordination
Phonetics–phonology interface

ABSTRACT

This paper takes a computational/experimental approach to investigating faithfulness in input–output phonological mappings. We seek to explain the results of a speech production experiment recently reported in Davidson (2010). In that experiment, native English speakers were asked to produce phonotactically unattested consonant clusters. We argue that modifications of the target consonant clusters are best understood by considering both unfaithful phonological mappings and imprecision in the speech production mechanism. To account for the pattern of unfaithful input–output mappings, we consider an extension of the P-map hypothesis (Steriade, 2008) to the production of phonotactically unattested target sequences. Predictions of the P-map for this data were established by a perception experiment in which participants were asked to discriminate between unattested consonant clusters, CC, and attested modifications, including epenthesis, prothesis, C_1 change and C_1 deletion. To evaluate the effects of motor noise on consonant cluster production, we constructed a computational model that allowed us to simulate consonant cluster productions under different levels of noise. Simulations reveal that, first, a large proportion of cases involving vocoid insertion, CəC, are better accounted for by noisy implementation of the target timing than by phonological epenthesis and, second, that differences in insertion patterns between stop-initial clusters and fricative-initial clusters are due to the internal temporal properties of stops and fricatives. Factoring these cases into the analysis isolates a pattern of unfaithful input–output mappings used to evaluate the P-map hypothesis. On the basis of considerable mismatches between patterns of perceptual similarity and unfaithful input–output mappings, we argue that the P-map theory of faithfulness is too restrictive to extend to the production of non-native speech.

1. Introduction

Bringing experimental data to bear on theoretical issues often requires teasing apart influences on performance in a specific task that are external to the grammar (see Goldrick, 2011 for review). In this paper, we focus on the task of producing non-native phonotactics—consonant clusters that are not attested in English—and develop computational tools for recovering the output of the phonological grammar from speech production data.

[*] Corresponding author at: MARCS Auditory Laboratories/School of Humanities and Languages, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia. Tel.: +61 2 9772 6275; fax: +61 2 9772 6688.
E-mail address: J.Shaw@uws.edu.au (J.A. Shaw).

0024-3841/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2011.03.003

Our primary theoretical question is the nature of faithfulness in Optimality Theory (Prince and Smolensky, 2004). Optimality Theory evaluates a set of candidate outputs with respect to two types of constraints, markedness constraints and faithfulness constraints. When high ranking markedness constraints prevent an input from surfacing faithfully, the internal ranking of faithfulness constraints can determine the optimal output, e.g., MAX ≫ DEP favors epenthesis mappings while DEP ≫ MAX favors mappings involving deletion (McCarthy and Prince, 1995).

Although, in principle, any ranking of faithfulness constraints may constitute a possible grammar, it has been argued that the set of attested languages do not take advantage of the full range of faithfulness rankings. This has been termed the "too many solutions problem". Steriade (2008) observes that proposed faithfulness rankings seem to maximize the perceptual similarity of input–output pairs. She achieves a formal account of this observation by projecting faithfulness rankings from a P-map (for "perceptual map"). The P-map is set of statements about the relative perceptual similarity of phonological forms. The grammar references these statements in ranking faithfulness constraints to maximize input–output similarity. By linking formal grammar to perceptual facts, the P-map offers a restrictive theory of faithfulness and a reply to the suggested problem of "too many solutions". Insofar as speakers of different languages share P-maps, i.e., judgments about the relative similarity of phonological forms, they are predicted to share faithfulness rankings and, hence, repair strategies in speech production.

Although the P-map was proposed to account for phonological alternations observed in adult speakers producing their native language, we extend the hypothesis to the production of non-native phonotactics. This extension builds on a body of research demonstrating influences of first language (L1) phonology and/or universal principles of markedness on cross-language speech production (e.g., Best and Tyler, 2007; Broselow et al., 1998; Eckman, 1977; Flege, 1995; Lado, 1957). By grounding faithfulness rankings in perceptual facts, the P-map hypothesis links perception and production yielding testable predictions for non-native phonotactics. When native English speakers are asked to produce consonant clusters that do not occur in English, they can make a variety of errors, or repairs of the target sequence (Davidson, 2006a, 2010). Our extension of the P-map predicts that the maximally harmonic output is the candidate that is most perceptually similar to the input consonant cluster. To explicate and evaluate this prediction, we conducted two experiments. First, a discrimination paradigm was used to determine the perceptual similarity of consonant clusters to minimally different forms (Davidson and Shaw, submitted for publication). Second, a speech production experiment was conducted to see whether the production patterns matched the perceptual facts (Davidson, 2010).

One way in which English speakers may deviate from target CC sequences is to produce them such that a short period of voicing surfaces between consonants. While determining the phonological status of this short period of voicing is non-trivial, it is crucial to evaluating our extension of the P-map hypothesis. The debate regarding the phonological nature of transitional vocoids has been taken up in the description of a number of languages, i.e., Berber (Dell and Elmedlaoui, 1996; Ridouane, 2008), Moroccan Arabic (Dell and Elmedlaoui, 2002; Gafos, 2002), Scots Gaelic (Bosch and de Jong, 1997), Hocank (Baertsch and Davis, 2009; Hall, 2006; Steriade, 1990) and Piro (Lin, 1997). In these languages, the data are rich enough to support phonological arguments for the status of voiced transitions. In non-native speech production experiments, the evidence bearing on the nature of inter-consonantal voiced transitions is often limited to phonetic data.

Whether or not a period of voicing is analyzed as a vowel or a transition between consonants affects conclusions drawn about the nature of the phonological grammar. In the case of inter-consonantal transitional vocoids, the nature of the voicing period is crucial to analyses of phonological epenthesis, and, consequently, the ranking of faithfulness constraints in OT. In order to navigate from raw speech production data to the output of the phonological grammar, this paper contributes a new computational tool. We adapt the probabilistic model of consonant cluster timing developed in Shaw et al. (2009) and Shaw and Gafos (2010) to explore patterns of inter-consonantal vocoids. The model allows us to predict the temporal properties of transitional vocoids that surface due to imprecise production of consonant clusters and evaluate those properties against the data. This gives us the ability to distinguish between the phonological outputs CəC and C^C, where '^' denotes a transitional vocoid. By factoring the effects of speech production imprecision into the analysis, we are able to isolate the data that bears most directly on our primary theoretical question, the perceptual basis of faithfulness.

The remainder of this paper is organized as follows. Section 2 reviews the results of a perception experiment. This experiment serves to establish the perceptual similarity of non-native consonant clusters to the space of possible repairs. Section 3 reports the results of a production study and evaluates whether, in line with P-map hypothesis, input–output mappings minimize perceptual distance. Section 4 describes a stochastic model of consonant cluster timing and two computational simulations. The first simulation renders consonant clusters at different levels of production noise. The results establish the pattern of inter-consonantal voicing expected to arise from temporal imprecision in speech production. The second simulation explores how the overall pattern of inter-consonantal voicing changes when a percentage of the simulated vocoids is generated from epenthetic vowels. We find that the best fit to the speech production data comes from the simulations based on a small percentage of epenthesis and a large percentage of vocoids generated from noise in consonantal timing. The simulation result allows us to adjust the overall insertion percentages to account for vocoids that are due to speech production noise, revealing a new pattern to be accounted for by the phonology. In section 5, we reconsider the P-map hypothesis in light of the simulation results. Section 6 briefly concludes.

2. Establishing perceptual similarity

In order to establish the predictions of the P-map hypothesis for English consonant clusters, we conducted a perceptual experiment. Following assumptions discussed by Fleischhacker (2005) we employed a discrimination task to assess perceptual similarity. Subjects were asked to discriminate between sequences of sounds attested in English and obstruent-obstruent and obstruent-nasal word-initial consonant clusters that are not possible in English, e.g. [fmatu], [gdase] (see also Berent et al., 2009; Berent et al., 2007; Kabak and Idsardi, 2007 for related perceptual discrimination experiments). If such sequences were to be adapted by English speakers, a variety of possible modifications would result in a phonologically licit word, including vowel insertion ([fəmatu]), vowel prothesis ([əfmatu]), C_1 deletion ([matu]), or C_1 change ([smatu]). By examining which of these modifications leads to the most confusion with the unattested consonant cluster in an AX discrimination task, we assess the perceptual similarity of potential input–output phonological mappings.

2.1. Method

2.1.1. Participants

Listeners included 38 participants recruited primarily through New York University classes and a posting on Craigslist in New York. They ranged in age from 19 to 42. A brief questionnaire was administered to all participants before beginning the experiment. Participants answered questions about language background including their native language and knowledge of other languages. All participants reported that English was their native language and no participants reported any history of speech or hearing disorders, or knowledge of a language that contains the consonant clusters under study, e.g. Hebrew, Russian, Polish, etc. The listeners were paid $10 for their participation. The data from one participant was discarded because he failed to respond to more than half of the trials.

2.1.2. Materials

The target materials for the AX discrimination trials consisted of non-words containing initial obstruent-obstruent and obstruent-nasal sequences and matching non-words with modifications hypothesized to potentially be perceptually confusable for English speakers. The 20 onset clusters under investigation can be divided into four groups based on the combination of the manners of their consonant sequences: fricative-nasal (FN: [fm], [sm], [zm], [vm]), fricative-stop (FS: [fp], [sp], [zb], [vb]), stop-stop (SS: [dg], [gd], [tk], [kt]), and stop-nasal (SN: [bm], [dm], [pm], [tm]). Each of the 20 consonant combinations were used to make stimuli of the form CCáCV. The diacritic over the [a] denotes stress. In the discrimination trials, each word containing a consonant cluster was paired with matching words for each of the following modifications: insertion (e.g., [tmáfa]/[təmáfa]), C_1 deletion (e.g., [tmáfa]/[máfa]), prothesis (e.g., [tmáfa]/[ətmáfa]), and C_1 change. For the C_1 change modification, there was more than one trial, since it is possible that listeners could misperceive the identity of the initial consonant in more than one way (e.g. [tmafa]/[dmafa], [tmafa]/[pmafa]). There was a total of 93 cluster-modification trials. A full list of all 'different' trials is given in Appendix A. The 'same' trials consisted of each word paired with itself (the same token repeated), both for cluster and modification stimuli.

The stimuli were recorded by a bilingual English/Russian speaker. The speaker was not a phonetician but did have linguistic training. Words were read from a printed list. In order to ensure that listeners were responding 'different' on the basis of the onset of the word only, the –áCV portion from one stimulus in the paradigm was spliced onto all of the rest of the stimuli. For example, all of the words that [tmafa] was paired with, including [mafa], [təmafa], [ətmafa], [dmafa], [pmafa], [bmafa], all shared the physically same [-afa]. The –áCV portion was always cut from zero crossings to avoid acoustic artifacts. Splicing was carried out straightforwardly by looking for obvious landmarks such as the end of a stop burst in C_1 position or the beginning of middle formant attenuation for the nasals in C_2 position. The acoustic properties of all the stimuli are reported in Davidson and Shaw (submitted for publication).

2.1.3. Procedure

The experiment was implemented in ePrime. The different trials contained both orders of presentation, so each participant heard half of the stimuli with the cluster word first (e.g. [tmáfa]/[máfa]) and half with the modification word first (e.g. [ədgáse]/[dgáse]). In light of past work showing different levels of discrimination sensitivity at short and long interstimulus intervals (Pisoni, 1973; Strange and Shafer, 2008; Werker and Logan, 1985), we also included interstimulus interval (ISI) as a variable in this experiment. Each participant heard all of the stimuli with ISIs of both 250 ms and 1500 ms, for a total of 300 trials for each participant. The stimuli were blocked for ISI, and participants were given a break between blocks. The ISI conditions were counterbalanced so half of the participants heard the short ISI first and half heard the long ISI first.

Participants were seated in individual small, quiet rooms containing PC computers and Sennheiser headphones. They were given these instructions: "In the following task, you will hear sound files presented as pairs of words. In some of the pairs, the sound files that you hear will be slightly different from one another, and others will be the same. Your task is to decide whether or not the sound files that are played to you are exactly the same. After hearing the second word, decide whether the two sound files are the same or different." Using the E-Prime button box, participants were told to press a button labeled "S" with the index finger if they thought the sound files were the same, and "D" with the middle finger if they thought the sound files were different. Participants were encouraged to answer quickly, and were told to choose either "S" or "D" even if they were not sure of the answer. As soon as the participant made a response, there was a 2500 ms pause and the next trial started. Before beginning the experiment, participants were first given 8 practice trials to familiarize themselves with the task. Feedback was not provided on any of the practice trials or test blocks.

2.2. Results

An analysis of variance was conducted to examine participants' performance on the AX discrimination trials. The within-subjects independent variables were interstimulus interval (ISI: 250 ms, 1500 ms), sequence type (FN, FS, SN, SS), and modification (insertion, prothesis, C_1 change, C_1 deletion). Subjects were included as a random factor. The dependent variable was d' (d prime), a sensitivity measure that computes how easily a listener can detect whether or not the signal—in this study, the different trials—is present (Green and Swets, 1966; Macmillan and Creelman, 2005). A d' of 0 indicates that listeners are responding at chance. The maximum d' value is 4.65, which indicates that listeners are scoring at ceiling on both different and same trials. Because accuracy scores (proportion correct on different trials) are more easily understood, these results are shown in Fig. 1. Results for d' are given in Table 1.

Results for the ANOVA showed a main effect of sequence [$F(3, 108) = 30.05$, $p < 0.001$, partial $\eta^2 = .455$] and modification [$F(3, 108) = 28.05$, $p < 0.001$, partial $\eta^2 = .438$], but no effect of ISI [$F(1, 36) < 1$]. The interaction between sequence and modification [$F(9, 324) = 26.05$, $p < 0.001$, partial $\eta^2 = .420$] was significant, but none of the two- or three-way interactions with ISI were significant. A Student–Newman–Keuls post-hoc test indicates that the main effect of sequence resulted from significant differences among all of the cluster types (> indicates significantly higher d' scores, $p < 0.05$): FN (mean $d' = 3.33$) > FS (3.11) > SN (2.72) > SS (2.43). The main effect of modification also showed significant differences among the four types: C_1 deletion (3.27) > insertion (3.10) > C_1 change (2.89) > prothesis (2.32).

To investigate the interaction between sequence and modification, separate ANOVAs were conducted for each sequence with modification as the independent variable. The effect of modification was significant for each sequence type, FN [$F(3, 108) = 23.42$, $p < 0.001$, partial $\eta^2 = .394$]; FS [$F(3, 108) = 68.97$, $p < 0.001$, partial $\eta^2 = .657$], SN [$F(3, 108) = 11.31$, $p < 0.001$, partial $\eta^2 = .239$], SS [$F(3, 108) = 8.17$, $p < 0.001$, partial $\eta^2 = .185$], but the patterns of accuracy differed across sequences. For FN sequences, a Student–Newman–Keuls post-hoc test shows the following pattern of accuracy results: C_1 deletion > C_1 change = insertion > prothesis. For FS sequences, the post-hoc test showed that C_1 deletion = insertion > C_1 change > prothesis, and that prothesis was significantly less accurate than all other modification types. For SN, the post-hoc test showed the pattern insertion > prothesis > C_1 deletion = C_1 change. Finally, the pattern for SS was C_1 deletion = C_1 change > prothesis = insertion.

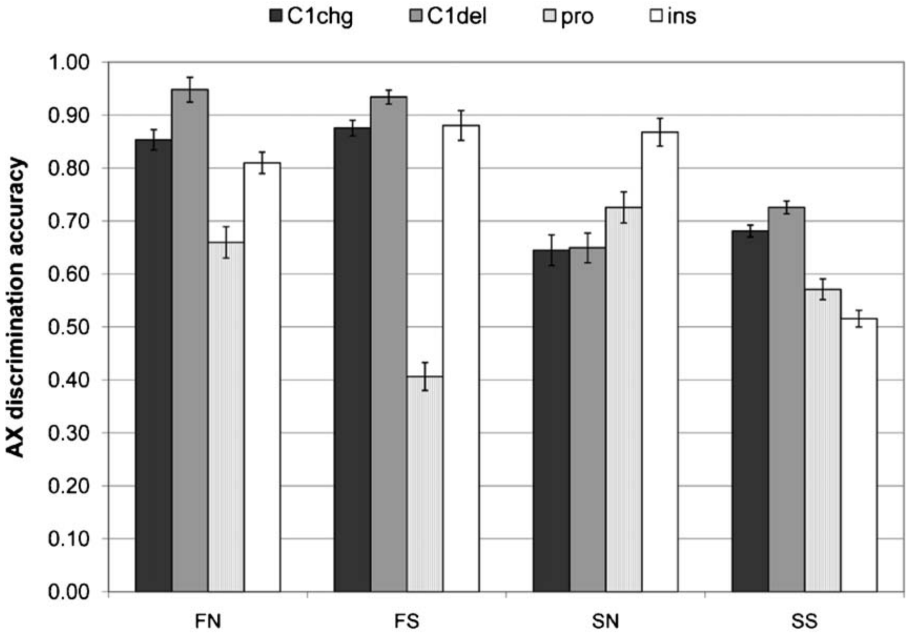

Fig. 1. Accuracy proportions for each type of modification, C_1 change (C1chg), C_1 deletion (c1del), prothesis (pro), insertion (ins) in the AX discrimination task. Results are divided by manner combination, fricative-nasal (FN), fricative-stop (FS), stop-nasal (SN), stop-stop (SS). Error bars indicate standard error.

Table 1
Table of d' scores for each sequence type and modification. The number in parentheses is standard error.

	FN	FS	SN	SS
C_1 change	3.327 (.124)	3.355 (.115)	2.286 (.144)	2.607 (.111)
C_1 deletion	4.017 (.104)	3.821 (.129)	2.446 (.190)	2.778 (.124)
Prothesis	2.663 (.163)	1.647 (.137)	2.751 (.137)	2.230 (.143)
Insertion	3.308 (.157)	3.613 (.146)	3.387 (.160)	2.095 (.144)

2.3. Summary

The results indicate that different consonant manner combinations give rise to different likely perceptual confusions. For both fricative-nasal and fricative-stop sequences, listeners were most likely to confuse the prothetic modification with the cluster. Stop-nasal sequences were confused with C_1 deletion and C_1 change most often, whereas insertion and prothesis were the most likely confusions for stop-stop sequences. These results suggest that the modification that is most perceptually similar to the unattested consonant sequence depends on the manner combination of the sequence. A discussion of how manner gives rise to different perceptual confusions is beyond the scope of this paper, but is discussed in detail in Davidson and Shaw (submitted for publication).

3. Patterns in non-native phonotactic production

The discrimination results determine the P-map's predictions for the production experiment. According to our extension of the P-map, productions of Russian consonant clusters by English speakers should include multiple types of errors. Furthermore, the most frequent error types should depend on the manner combination of the clusters. Unfaithful productions should involve prothesis for fricative-initial clusters, epenthesis or prothesis for stop-stop clusters and C_1 change or C_1 deletion for stop-nasal clusters. In this section, we turn to the speech production data to test these predictions.

Aspects of the speech production experiment described below were first reported as part of a larger study in Davidson (2010). Here we summarize the data most relevant to evaluating the P-map hypothesis, the results from the English speakers in the text + audio condition. In this condition, native English speakers saw an orthographic representation of the target word concurrent with hearing the auditory stimulus twice. After hearing the auditory stimulus, participants attempted to produce the target sequence as they heard it. The target words we focus on here contained the same type of obstruent-initial onset clusters as those in the perception experiment reported in section 2: fricative-stop (FS, e.g. [vbagu]), fricative-nasal (FN, e.g. [znagi]), stop-stop (SS, e.g. [bdava]), and stop-nasal (SN, e.g. [gmalo]). The production experiment also contained stop-fricative and fricative-fricative onsets, but those are not reported on here. In addition to the stimuli with cluster onsets, speakers also produced words containing an initial CəC sequence, i.e. [vəbagu], [zənagi], [bədava], [gəmalo]. The stimuli were recorded by a native English/Russian bilingual. Further details about the stimuli can be found in Davidson (2010).

All of the participants' responses were analyzed by repeatedly listening to the files and examining the spectrograms of the utterances in Praat to determine what, if any, error had been produced. The data were coded by a phonetically trained research assistant who was blind to the purpose of the experiment. Four repairs accounted for about 95% of the data: insertion, C_1 deletion, C_1 change, and prothesis. A token was coded for insertion if there was either a period of voicing after frication or after a stop burst with formant structure containing a visible second formant that ended with abrupt lowering of intensity at the onset of the second stop or nasal. If a token was produced with no evidence of C_1 (no stop burst or no frication noise), it was coded as C_1 deletion. C_1 change indicated that a speaker produced a consonant differing in manner, place, and/or voicing than was produced by the English/Russian bilingual speaker who recorded the stimuli. A token was coded for prothesis if the speaker uttered a vocalic portion containing formant structure before the stop silence or fricative noise. If no errors of these types occurred, the token was labeled as 'correct'. Example spectrograms of these repair types are presented in Appendix B.

Error results for the production of obstruent-stop and obstruent-nasal clusters are shown in Fig. 2. This graph illustrates the types of modifications English speakers produced when they failed to accurately produce the obstruent initial onset sequences. For each sequence type, insertion is a significantly more frequent modification than C_1 deletion, C_1 change, or prothesis. As for overall accuracy, speakers were significantly less accurate on the stop-initial sequences than on the fricative-initial ones, but there were no differences between SS and SN or FS and FN.

These findings demonstrate that English speakers make predominately the same error, vocoid insertion, for all unattested consonant clusters, regardless of the manner combination of the consonants. This does not align with the predictions of the P-map hypothesis. The results of the perception study indicated that insertion (as well as prothesis) should be produced for stop-stop clusters, but that for other manner combinations, different modifications should be produced. Fricative-initial sequences should be modified primarily by prothesis and stop-nasal sequences by C_1 change or C_1 deletion. Therefore, modifications of the target consonant clusters do not follow patterns of perceptual similarity.

If the P-map does not account for patterns of consonant cluster modification in production, then what does? We focus our account on two aspects of the pattern. The first is the predominance of vocoid insertion as a modification of unattested consonant clusters. The second is the different percentages of vocoid insertion observed for fricative-initial vs. stop-initial clusters. We account for these aspects of the data by considering the possibility that some of the insertion errors coded in the production data arise from imprecision in speech production.

4. Temporal imprecision in speech production

In the production experiment, subjects were asked to produce consonant clusters that do not appear in English. Since phonotactically illicit sequences are unfamiliar and unpracticed, it is possible that English speakers lack the motor precision

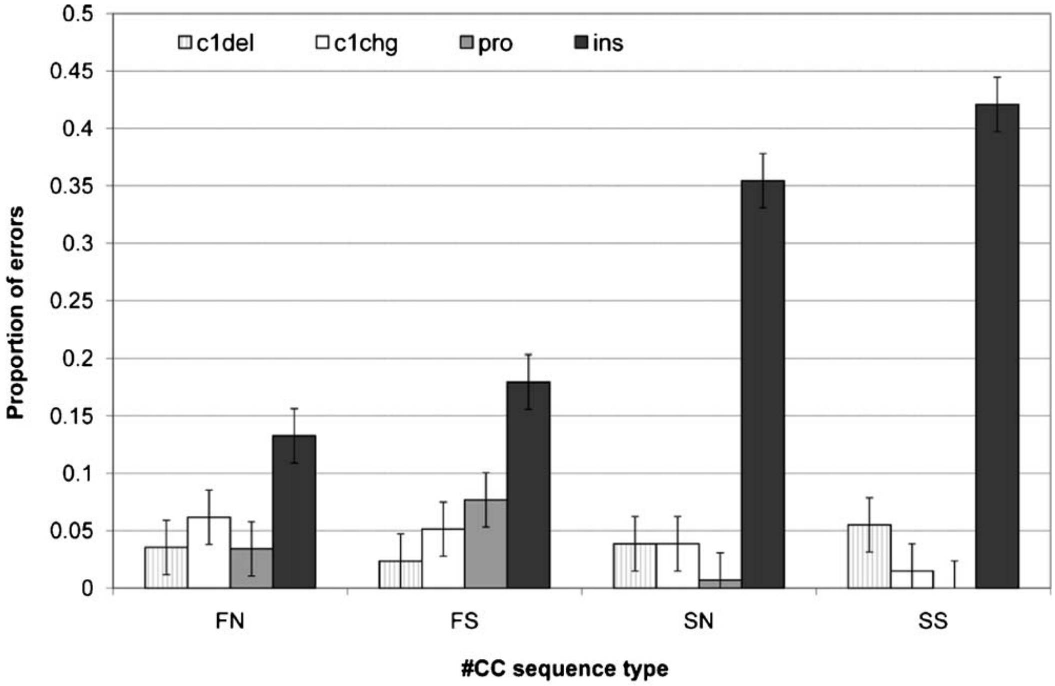

Fig. 2. Error proportions for each manner combination, fricative-nasal (FN), FS, SN, SS in the cluster production task. Error bars indicate standard error.

to implement them without error. Motor error, defined as deviation of articulatory movements from a temporal plan, may be independent of whether the phonological grammar renders faithful mappings of target sequences. Nevertheless, motor error can cause brief periods of voicing in the acoustic signal. How do these periods of voicing differ from schwa? To establish predictions, we constructed a computational model and simulated CC timing under different levels of noise. The simulated data shows how timing variability affects insertion percentage, vocoid duration and vocoid variability. After establishing these patterns, we return to the production data to test them.

4.1. Computational model

The model is structured to simulate the transition between consonants in initial clusters, e.g., #CCV-. We begin by assuming fully faithful mappings between the phonological input, which we take to be the target phonological string, and the surface phonological form. The model then renders the timing between consonant clusters under different levels of Gaussian distributed noise. At low levels of noise, consonant clusters are generated without voicing between consonants, replicating the Russian-like stimuli presented in the experiments. This was ensured by setting the onset of voicing following C_1 to occur at a greater interval from the release of C_1 than the achievement of target of C_2. When C_2 constriction is achieved before the onset of voicing following C_1, then there can be no voicing between consonants in the acoustics. Under perturbation by noise, however, the target timing relations may be actuated such that voicing surfaces in the acoustic record. These two scenarios are schematized in Fig. 3. The left side of the figure provides a schematic depiction of the target timing. The right side of the figure shows a case of vocoid "insertion" due to noisy actuation of the target coordination relations. By simulating consonant cluster production at different levels of noise, we can evaluate whether the levels of noise required to yield insertion percentages characteristic of the data also capture the temporal properties of voiced vocoids.

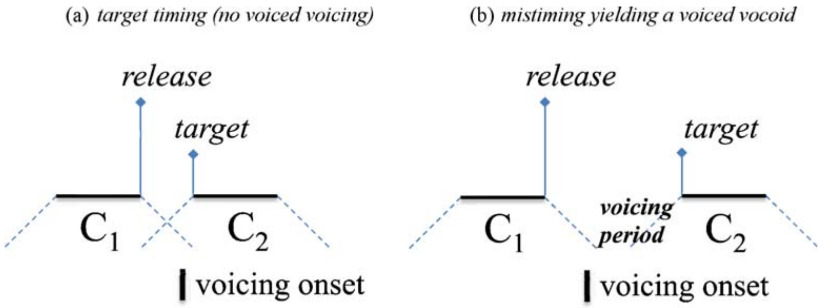

Fig. 3. CC timing with interconsonantal voicing (right) and without voicing (left). (a) Target timing (no voiced voicing). (b) Mistiming yielding a voiced vocoid.

4.1.1. Model parameters

The approach taken here is related to earlier modeling work on consonant cluster timing reported in, Shaw and Gafos (2010), Shaw et al. (2009, submitted for publication), Gafos (2002), and Browman and Goldstein (1990). In this case, we are interested in how noise in the timing of consonant oral constrictions affects the appearance of a voiced vocoid between adjacent consonants. To address this question, we modeled the temporal landmarks that define the interval between consonantal plateaus, i.e. the constriction phase of the consonants. Relevant landmarks are the release of the first consonant, C_1 in $\#C_1C_2$ sequences, and the achievement of target of the second consonant, C_2 in $\#C_1C_2$ sequences. We assume that the timing between C_1 and C_2 is a property of the structural relation between consonants in a word-initial cluster and that it remains constant across stop-initial and fricative-initial clusters.

We also modeled the onset of voicing following C_1. We assume that the timing between C_1 release and voicing onset is a property of C_1. Since one focus of the modeling is on the difference between stop-initial and fricative-initial clusters, the model formalizes internal temporal differences between these types of segments. Below, we substantiate the claim that the interval from the release of C_1 to the onset of voicing is longer for fricatives than for stops. As model simulations will illustrate, this difference in the internal structure of stops and fricatives is central to determining insertion patterns for fricative-initial vs. stop-initial consonant clusters.

All landmarks were generated from stochastic versions of local timing relations. The specific algorithm is summarized in Fig. 4. Landmark generation proceeds by first selecting the achievement of target of C_2 from a normal Gaussian distribution. The mean duration between the plateaus of the consonants was determined by a constant, k^{ipi}, encoding the inter-plateau interval (ipi). The release of C_1 was generated by subtracting k^{ipi} and adding a noise term, ε^{ipi}. The onset of voicing following C_1 was generated by adding a constant, k^{vot}, encoding voice onset time (vot), to the release of C_1 and adding another noise term, ε^{vot}. As shown schematically in Fig. 4, the voicing onset for stops was determined by a shorter value of the k^{vot} constant than for fricatives.

To summarize, the model specified above treats the timing between oral constrictions equivalently in both stop-initial and fricative-initial clusters. However, the internal temporal organization of stops and fricatives is differentiated using the k^{vot} constant. Specifically, the value of k^{vot} is set to a higher value for fricatives than for stops. This temporal difference between stops and fricatives plays an important role in conditioning insertion patterns for fricative-initial vs. stop-initial consonant clusters. Before reporting simulation results, we first provide independent motivation for differentiating the value of the k^{vot} constant according to C_1 continuancy.

4.1.2. Differences in the temporal structure of stops and fricatives

The interval extended from the release of C_1 to the onset of voicing has been studied extensively in stops where it is acoustically salient and encodes phonological voicing distinctions (Cho and Ladefoged, 1999; Lisker and Abramson, 1964, 1967). Much less is known about this temporal interval in fricatives. In fricatives the release from constriction is difficult to decipher from the acoustic signal and, as far as we know, does not distinguish phonological contrast. For this reason, we provide some brief background on fricative releases and substantiate suggestions in the literature with illustrative articulatory phonetic data.

The acoustics of fricative offsets generally involve gradual attenuation of the level of fricative noise. Docherty (1992:118–119) suggests that this attenuation may correspond to the gradual widening of the primary area of constriction. Consistent with this suggestion, Catford (1977:254) notes that the noise associated with a fricative can continue for some time after the release of the oral constriction. To exemplify these suggestions, Fig. 5 contrasts fricative and stop releases. The data are drawn from the publicly available Wisconsin X-ray Microbeam Speech Production Database (Westbury, 1994). The figure shows articulatory and acoustic data of productions of [d], left, and [z], right, in matched intervocalic contexts, [ə_a]. The stimuli presented to the speaker were *uhda* and *uhza*, orthographic strings intended to prompt, respectively, productions of [əda] and [əza]. The data shown here were produced by speaker 11 as a part of Task 16, which includes the production of

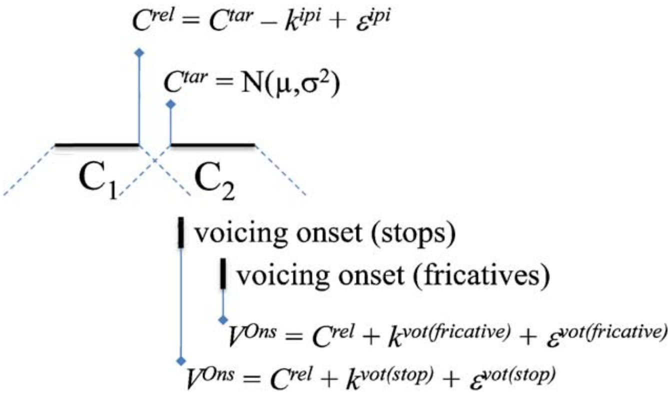
$C^{rel} = C^{tar} - k^{ipi} + \varepsilon^{ipi}$
$C^{tar} = N(\mu,\sigma^2)$
C_1
C_2
voicing onset (stops)
voicing onset (fricatives)
$V^{Ons} = C^{rel} + k^{vot(fricative)} + \varepsilon^{vot(fricative)}$
$V^{Ons} = C^{rel} + k^{vot(stop)} + \varepsilon^{vot(stop)}$

Fig. 4. Summary of the algorithm generating landmarks.

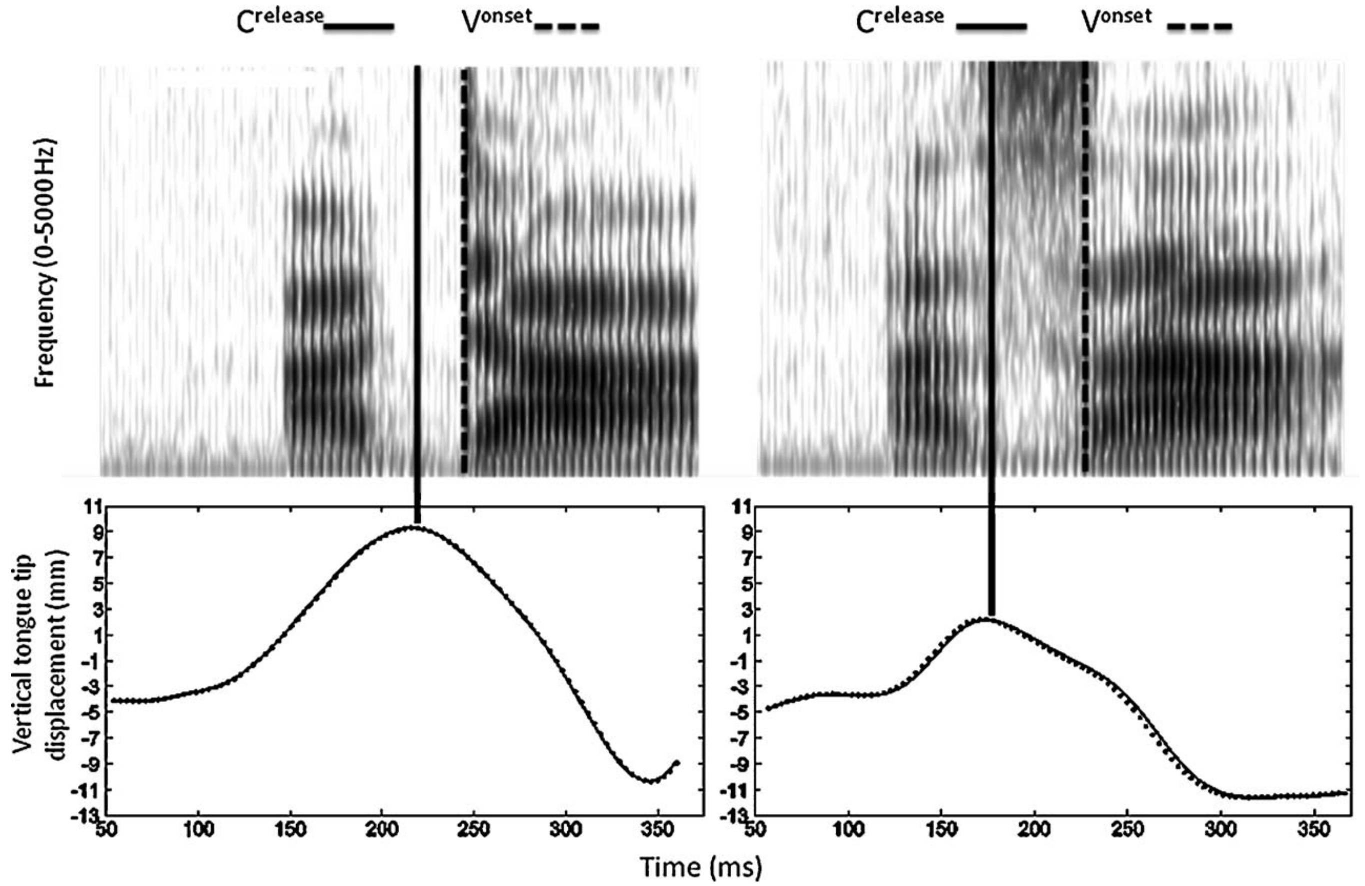

Fig. 5. X-ray microbeam data (bottom panels) and corresponding spectrograms (top) showing stop (left) and fricative (right) releases. The bottom panel shows the position of the tongue tip sensor (T1) in the vertical dimension as a function of time. The left panel shows the production of /d/ and the right panel shows the production of /z/.

multiple consonants in the same intervocalic context. The bottom panel of the figure shows movement of the tongue tip (T1 sensor) in the vertical dimension as a function of time. The top panel shows the corresponding spectrogram. Consistent with the descriptions of fricative releases reviewed above, the right side of the figure shows that frication noise in the spectrogram continues as the release of the tongue tip gradually widens the aperture of constriction. For comparison, the bottom left panel shows tongue tip release for a stop. The relevant point for setting the k^{vot} parameter in the model is that the interval from consonantal release, solid black line, to the onset of voicing, dotted black line, is longer for fricatives than for stops.

The point of Fig. 5 is that it illustrates a temporal difference between stops and fricatives. The interval from release of a consonant constriction to the onset of voicing is longer in fricatives than in stops. This difference, shown schematically in Fig. 4, was encoded in the model by ensuring that the value of $k^{vot(fricative)}$ is greater than $k^{vot(stop)}$. As long as this inequality holds up, the model makes predictions also expressible in terms of inequalities. First, vocoids will surface at a lower level of variability in SS clusters than in FS clusters. Second, at any fixed level of variability, there will be more inserted vocoids in SS clusters than in FS clusters. Third, inserted vocoids will be, on average, longer in SS than in FS clusters. These statements of inequality are predictions of the model regardless of the specific values of the parameters. They follow from the internal temporal differences of stops and fricatives. Nevertheless, statements of inequality may lack the precision to differentiate competing hypotheses about phonological structure (Shaw and Gafos, 2010). Model predictions are made more precise by specifying values of model parameters and simulating data. Measurements of the experimental data can then be directly compared to the simulated data.

4.1.3. Simulations

The model described above was used to simulate consonant cluster timing under gradually increasing levels of noise. Although each landmark in the model is associated with an error term, only the error term linked to CC timing, ε^{ipi}, was manipulated. In general, model parameters were estimated to represent the phonetic parameters underlying the data. To instantiate our assumption that FS and SS clusters share the same phonological structure, the same value for k^{ipi}, 20 ms, was used for both manner combinations. The error term associated with this constant, ε^{ipi}, was manipulated across runs of the simulation. This parameter was drawn from a normal Gaussian distribution with a mean of 0 ms and a standard deviation ranging across runs from 5 ms to 105 ms.

The constant specifying the onset of voicing following C_1 was set according to manner. To ensure that voicing does not surface between consonants at low levels of noise, k^{vot} was set to be greater than k^{ipi} (as schematized in Fig. 4). For stops, k^{vot} was set to 30 ms. For fricatives, k^{vot} was set to 60 ms. This difference is on the order of magnitude found in the X-ray microbeam data (Fig. 5). The error terms associated with these landmarks, the onset of voicing for fricatives and stops, as well as the error term associated with the achievement of target of C2, were held constant at 20 ms.

Simulations were conducted in MATLAB. On each run, 5000 stop-initial and 5000 fricative-initial consonant clusters were simulated. On the first run, the error term associated with the release of C_1 was set to 5 ms and on each of 100 subsequent runs this value was increased by one. This provided a range of variability from 5 ms to 105 ms. We report on each of the 100 runs (10,000 consonant clusters per run) of the simulation. On each run, the number of consonant clusters simulated with a voiced vocoid was tabulated and the duration of the voiced vocoid was measured. A voiced vocoid was counted if there was more than 6 ms (approximately one period) between the onset of voicing following C_1 and the achievement of target of C_2. The duration of the vocoid was determined by measuring the interval between the onset of voicing landmark and the achievement of target of C_2.

Fig. 6 shows the simulation results. The top panel summarizes the proportion of insertion trials, the middle panel shows the mean duration of the voiced vocoids, and the bottom panel shows the standard deviation of the voiced vocoids. As ε^{ipi} increases, each of these measurements tends to increase. In some cases, however, there is a local decrease. For example, in the step from $\varepsilon^{ipi} = 38$ to $\varepsilon^{ipi} = 39$, the insertion percentage (top panel) decreases temporarily before increasing again at step 40. These dips illustrate a dependency between the three measurements in Fig. 6. With the small dip in insertion percentage comes, as well, small decreases in vocoid duration (middle panel) and vocoid variability (bottom panel). The phonological structure encoded in the model, [CC], thus predicts a reciprocal relation between these measurements. Even as the proportion of insertion changes (as ε^{ipi} is scaled), the relationship between insertion percentage, vocoid duration and vocoid standard deviation remains constant. Thus, there is an invariant relation between measurements in the simulated data. This relation characterizes the CC structure of the model. As such, it can be used to diagnose this structure in the phonetic data.

The simulated data can be compared to the experimental data in a number of different ways. We emphasize here the importance of considering the relationship between measurements. The simulations show that any one measurement ranges widely as a function of variability. Nevertheless, the relation between measurements remains a constant indicator of phonological structure. In comparing the simulated data to the experimental data, we will therefore be looking across the different measurements. Since the focus of our simulation is on the effects of C_1 continuancy and of noise in consonant cluster timing, we, for the purposes of model evaluation, abstract away from the nasality of C_2. The simulated data will therefore be evaluated against the range of experimental data produced with stop-initial (including both SS and SN) and fricative-initial (including both FS and FN) clusters. We proceed by identifying the levels of variability in consonant timing (value of ε^{ipi}) that best account for the proportion of insertion found in the data. We then examine whether the simulated data capture vocoid duration at these same levels of variability. By looking at insertion percentage and vocoid duration

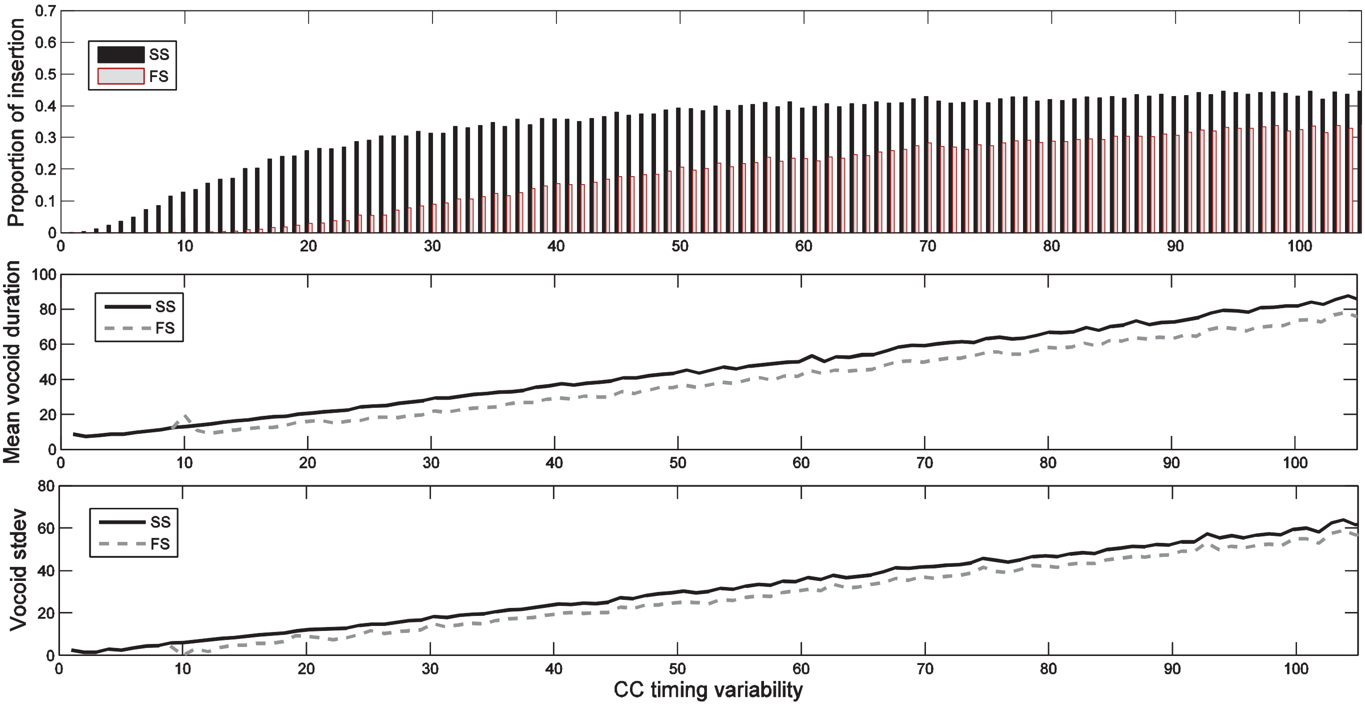

Fig. 6. Simulation results generated by a model of CC structure at different levels of timing variability (*x*-axis). The top panel shows the proportion of inserted vocoids, the middle panel shows the mean vocoid duration and the bottom panel shows the standard deviation of inserted vocoids.

Table 2
Vocoid duration by stimulus type and C_1 manner {fricative, stop}. The "C^C" column shows the mean duration of transitional vocoids produced in CC targets and, in parenthesis, the standard deviation of the transitional vocoid. The "CəC" column shows the mean duration of schwa in CəC targets. The standard deviation of the vowel is given in parenthesis. All values are in milliseconds.

	C^C		CəC	
	Mean	SD	Mean	SD
Fricative-initial	45	18	60	20
Stop-initial	40	16	52	21

within the same range of ε^{ipi} we can effectively evaluate whether the relation between these parameters is shared across the simulated and experimental data.

Before proceeding with an evaluation of the model, we briefly review the experimental data. Insertion percentage, as shown in Fig. 2, was greater for stop-initial clusters, ranging from 36% to 42% (36% for SN and 42% for SS), than for fricative-initial clusters, ranging from 13% to 17% (13% for FN and 17% for FS). As for vocoid duration, Davidson (2010) reports that periods of inter-consonantal voicing were significantly longer in CəC targets than in CC targets. Here, we are interested not only in this overall difference but also in differences between stop-initial and fricative-initial clusters, and in the standard deviation of the vocoids. Table 2 provides these measurements. For both stop-initial clusters and fricative-initial clusters, the vocoid in CəC targets was produced with longer duration than the vocoid in CC targets. Another important point is that both schwa (ə), and inserted vocoids (^), are longer after fricatives than after stops.

We now evaluate how closely the simulated data captures relationships between measurements of the experimental data. First of all, we note that, regardless of the level of timing variability, stop-initial clusters had higher rates of insertion than fricative-initial clusters. This is the same qualitative pattern found in the data. As discussed above, this follows from differences in the internal temporal structure of stops and fricatives. Second, we note that there are levels of variability at which the simulated data provide a precise quantitative match to the measurements of the experimental data. As described above, however, we are interesting in evaluating the relation between measurements. Our evaluation strategy is to fix the range of variability at which the model captures insertion percentage and to look at the other measurements to see if they match the data within the same range.

At step 37 on the *x*-axis, the simulation matches the bottom of the insertion ranges found in the data for both stop-initial, 36%, and fricative-initial, 13%, clusters. The upper range of the insertion proportions are best captured at variability step 44. At this step on the *x*-axis, Fig. 6 shows 17% for FS and 39% for SS. Thus, for insertion percentage, the model comes closest to the data between steps 37 and 44. We now focus on the duration of the vocoids within this range. Between steps 37 and 44 on the *x*-axis, the model produces vocoids between stop-initial clusters with a mean duration ranging between 33 ms and 38 ms. The data, however, has a mean duration of 40 ms. Thus, the model slightly underestimates vocoids duration in SC sequences. For fricative-initial clusters, the problem is worse. In the relevant range, the model predicts vocoids in fricative-initial clusters between 27 ms and 30 ms, but the mean duration of these vocoids in the data is 45 ms. Another problem is that, in the simulated data, the transitional vocoid following stops is longer and more variable (higher standard deviation) than the transitional vocoid following fricatives. In the experimental data, we see the opposite pattern. The duration of the transitional vocoid was longer in FS sequences than in SS sequences.

In sum, we have argued that the relation between insertion percentage and vocoid duration characterize CC structure. The model was able to capture insertion patterns and vocoid duration, separately, but not the relation between these two measurements. Within the range of variability under which the model captures insertion patterns, the simulated vocoids are systematically shorter than in the experimental data the relative duration of vocoids in FS and SS clusters goes in the opposite direction of the data.

4.2. Phonological variation

We have thus far explored two different explanations for the observed pattern of consonant cluster modification by English speakers. The first explanation is that the pattern of modification is dictated by the phonological grammar. The second explanation is that the error pattern is due to temporal imprecision in speech production. Neither of these approaches, by themselves, account for the full range of data. The phonological explanation we pursued predicts, on the basis of the perception results, a smaller proportion of vocoid insertion than was observed. The noisy production hypothesis yields the right percentages of insertion but, in doing so, also makes incorrect predictions about vocoid duration patterns.

We now turn to a third hypothesis that makes use of both unfaithful phonological mappings and noisy implementation of target clusters. The central idea is that the data involves a mix of phonological epenthesis and instances of vocoid insertion attributable to inaccurate achievement of temporal targets. Thus, there is variation in the phonological component of the grammar. A target consonant cluster, /CC/, maps sometimes to [CəC] and sometimes to [CC], which may be produced with or without a transitional vocoid. We find that both the insertion percentages and the vocoid duration facts can be accounted for by considering in tandem how mechanical noise and phonological variation shape the data. The modeling paradigm developed in the previous section allows us to test this "mixed" hypothesis.

To generate quantitative predictions, we conducted a new simulation, this time including a mix of CC and CəC forms. As with the CC forms, CəC forms were simulated from probabilistic versions of local timing relations. The duration of the interval extending from the onset of voicing following C_1 to the achievement of target of C_2 was determined, for CəC forms, by vowel duration. Vowel duration was sampled from a normal distribution matching the data (reported in Table 2, section 4.1). For fricative-initial forms, FəC, the vowel duration distribution had a mean of 60 ms and a standard deviation of 20. For stop-initial forms, SəC, the vowel duration had a mean of 52 ms and a standard deviation of 21.

We explored different combinations of CC and CəC forms; here we report the combination that best fits the data. In this simulation, the model generated 90% CC forms and 10% CəC forms. On each run, 10,000 total forms (5000 fricative-initial and 5000 stop-initial) were simulated. Of this total, 9000 (4500 FC and 4500 SC) were simulated from the CC structure and 1000 (500 FəC and 500 SəC) were simulated from the CəC structure. For CC forms, all model parameters were identical to the previous simulation. As before, the level of variability in the CC forms (but not in the CəC forms) was increased on each run of the simulation from 5 ms to 105 ms. Therefore, the only difference between this simulation and the simulation reported in section 4.1 is that in this simulation 10% of the forms had CəC structure. The results were analyzed to determine the percentage of trials on which a vocoid surfaced between target consonants, as well as the mean and standard deviation of vocoid duration. This analysis was blind to the structure (CC or CəC) that gave rise to each simulated form. Thus, the results illustrate a pattern derived from calculating statistics over a non-uniform set of phonological structures (both CC and CəC).

Fig. 7 shows the results of the mixed 90% CC/10% CəC simulation. The format of the figure is identical to Fig. 6. The top panel shows that there is more insertion in stop-initial clusters than in fricative-initial clusters. This replicates the result of the previous simulation. Because vowels arising from CəC are counted as insertion, the levels of insertion percentage in this simulation reach the data at lower levels of ε^{ipi}. That is, less variability is required to match the proportion of inserted vocoids found in the speech production experiment. The middle panel of Fig. 7 shows vocoid duration. Changes in vocoid duration as ε^{ipi} is scaled are more complex than in the CC only model. The average duration of the vocoid is longer than the CC only simulation, particularly at low levels of variability. This is because 10% of the inserted vocoids are actually vowels and, at low levels of variability, vowels are longer than non-vowel transitional vocoids. Another key difference between this simulation and the CC only simulation reported above is that the effect of C1 manner on vocoid duration is reversed. At low levels of variability (between steps 1 and 38 on the *x*-axis) the vocoid measured in FS clusters is longer than in SS clusters (in the last simulation, CC-only, the vocoid in SS clusters was longer than the vocoid in FS clusters at all levels of variability). Each of the above changes in model output arises from incorporating a small percentage of true vowels into the simulation, and each of these changes shifts the overall pattern of the simulations in the direction of the data.

Most importantly, the data simulated from a combination of CC and CəC matches multiple measurements in the experimental data with the same parameter values. Between steps 18 and 26 on the *x*-axis, the model captures the fricative-initial insertion percentages (13–17%). Between steps 19 and 30, the model captures the stop-initial insertion percentages (36–42%). Thus, in the range from 19 to 26, the model-simulated insertion percentages match the data for both stop-initial and fricative-initial clusters. At this range of variability, 19–26 on the *x*-axis, the mean duration of the inter-consonantal

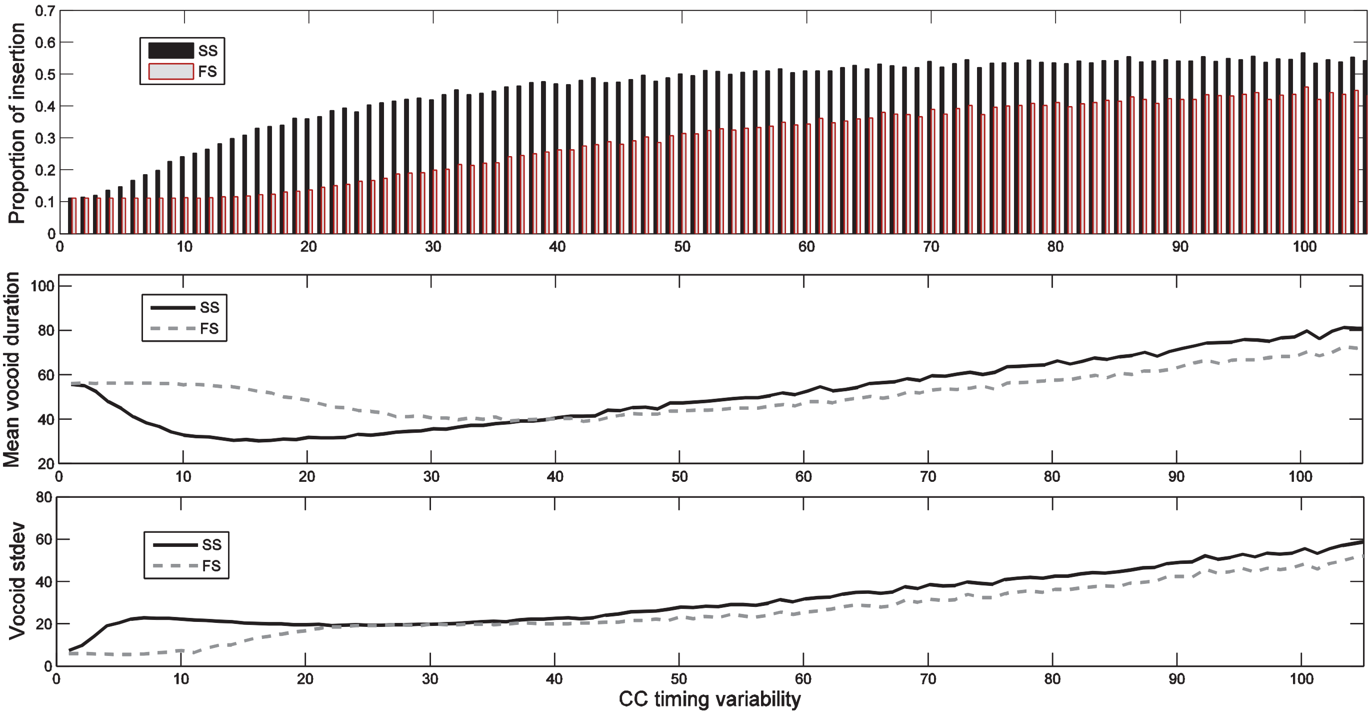

Fig. 7. Results of simulation generating 90% CC clusters and 10% CəC sequences at different levels of timing variability (*x*-axis). The top panel shows the proportion of inserted vocoids, the middle panel shows mean vocoid duration and the bottom panel shows the standard deviation of the inserted vocoid.

vocoid in fricative-initial clusters is between 43 ms and 49 ms. This prediction encompasses the actual mean duration in the speech production data, 45 ms. Moving now to vocoid standard deviation, at the same range of variability, from steps 19 to 26, the simulated data matches the experimental data. In the experimental data, the standard deviation of inserted vocoids was 18 ms in fricative-initial targets and 16 ms in stop-initial targets. The bottom panel of Fig. 7 shows that, in the relevant variability range, the model predicts a standard deviation between 18 and 21 ms for vocoids in fricative-initial targets and a standard deviation of 16–20 ms for vocoids in stop-initial targets. Crucially, therefore, the simulated data captures the relation between insertion percentage, vocoid duration, and vocoid standard deviation that is found in the data.

The hybrid model composed of a mix of CC and CəC sequences improves over the CC only model in multiple ways. It captures the effect of C1 continuancy on vocoid duration and as well as relationships between insertion percentage, vocoid duration and vocoid variability. Nevertheless, there is one measurement on which the simulated data and experimental data still do not match (in the range of 19–26 on *x*-axis). For stop-initial clusters, the model still underestimates the duration of the transitional vocoid. In the speech production data, this interval is 40 ms. In the target variability range, the model predicts the mean vocoid duration following stops to be somewhat lower, between 31 and 33 ms. This cannot be easily remedied. While increasing the number of CəC sequences for stop-initial clusters would push the mean vocoid duration towards the data it would also push insertion percentage away from the data. One possibility is that, for stop-initial sequences, talkers adjust the timing between consonants at a phonological level. That is, consonant clusters are pulled apart in time, not because of low level motor error but as a way to avoid phonotactic violation (Davidson, 2006b; Gafos, 2002). In the context of the current modeling paradigm this could be tested by manipulating the k^{ipi} parameter. However, we leave this matter for future research.

To summarize, the mixed (90% CC/10% CəC) simulation improves on previous simulation results. While the CC-only simulation captured aspects of the data individually, the mixed simulation captured the relation between insertion percentage, vocoid duration and vocoid stability. The mixed hypothesis also improves on the CC-only model by capturing how C_1 continuancy affects the duration of following vocoids. Vowels in unstressed syllables are longer when they follow fricatives. By incorporating a small percentage of true vowels into the simulation, the overall pattern shifts in the direction of the data.

Overall, the simulation results demonstrate, firstly, that noisy implementation of target consonant clusters can go a long way towards accounting for both the preponderance of vocoid insertion in the data and the relationship between C_1 continuancy and vocoid insertion percentage. Secondly, by holding the model accountable for the relation between insertion percentages and vocoid duration/stability, we found that the data are better characterized as the product of multiple target structures, both CC and CəC, than as the consequence of noisy implementation of CC targets only. Specifically, the simulation based on just CC targets failed to capture vocoid duration patterns. When a small number of targets with epenthetic vowels, CəC targets, were added, the simulation successfully captured vocoid patterns as well as insertion percentages.

A major consequence of the modeling results is that they force us to adjust the data with which we evaluate phonological theory. The modeling results indicate that there are some cases of true epenthesis mixed in with cases of noisy implementation of target CC. The mixture of CC and CəC targets required to match the data constitutes a specification of the phonological output, the level of representation that the P-map is hypothesized to explain.

5. Back to input–output mappings

We take the simulation results to indicate that phonological epenthesis accounts for some cases of insertion coded in the speech production experiment. Having established a method for distilling cases of epenthesis from insertion, we are now in a better position to evaluate grammar on the basis of speech production data. We have already established that our extension of the P-map hypothesis fails to account for the complete pattern of insertion. However, we can now formulate a more appropriate question: Do the rankings projected from the P-map account for the subset of insertion cases attributable to epenthesis?

As the modeling results demonstrate, the majority of the insertion cases can be accounted for by noisy implementation. Setting these cases side, we are left with the 10% of the insertion modifications involving epenthesis. How does epenthesis compare with other unfaithful input–output mappings? For fricative-stop clusters, epenthesis in 10% of trials can be compared to prothesis in 8% of trials, C_1 change in 5%, and C_1 deletion in 2%. This still equates to more epenthesis than any other individual input–output change. After adjusting for noise in the production of stop-stop clusters, we are left with 10% epenthesis, 6% C_1 deletion and 1% C_1 change. While it remains the case that, for English speakers, epenthesis is the dominant process involved in mapping unattested consonant clusters to less marked phonological forms, the percentages of epenthesis are much smaller than the percentages of insertion.

The modeling results also allow us to adjust our view of the relative markedness of SS and FS clusters. In the raw data, fricative-initial clusters were more accurate, overall, than stop-initial clusters. In section 2, we reported the percentage of errors on each manner combination. There were more errors on stop-stop clusters than on fricative-stop clusters (67% accuracy for fricative-stop clusters vs. 51% accuracy for stop-stop clusters). Parceling out errors due to imprecise production reveals a more nuanced picture. The grammar produces more faithful mappings for stop-stop clusters than for fricative-stop clusters (83% faithful mappings for stop-stop clusters vs. 75% faithful mappings for fricative-stop clusters). Thus, there is a mismatch between accuracy patterns (higher accuracy on FS than SS) and faithful phonological mappings (more faithful mappings for SS than for FS). Stop-stop clusters are more likely to be rendered faithfully than fricative-stop clusters but also more likely to be implemented with a transitional vocoid than fricative-stop clusters. Our modeling paradigm provides a grammar-external explanation for this second fact. Explaining the first requires a deeper probe into the content of the phonological grammar. The data suggest that SS clusters are less marked than FS clusters.

One reason why unfaithful mappings for FS outnumber those for SS clusters is that, while both clusters are subject to epenthesis, FS clusters (and not SS clusters) are also subject to prothesis on some trials. This fact reconnects us to the role of perception in phonological patterns. We grounded our P-map predictions for non-native consonant clusters in the discrimination results of our perception experiment. According to those results, our extension of the P-map predicts that prothesis should be the predominant unfaithful mapping for fricative-initial consonant clusters. Contrary to this prediction, cases of epenthetic mappings outnumbered prothetic mappings for all manner combinations, including fricative-initial clusters. The prevalence of epenthesis across manner combinations regardless of the perceptual similarity of input–output forms speaks against our extension of the P-map hypothesis. Moreover, this result holds up even after cases of non-epenthesis voicing insertion are cleaned from the data. While the results are inconsistent with the P-map as determinant of unfaithful mappings, we leave open the possibility that perception may play some other role in shaping phonological patterns. In particular, we have not provided any explanation for the fact that, although small in number, prothetic mappings are attested for fricative-initial clusters (but not for stop-initial clusters). It may be possible to formulate a weaker role for perceptual similarity in input–output mappings. Such an account would also have to explain why prothetic mappings did not occur at all for SS clusters despite the high perceptual similarity of SS∼əSS pairs. It seems, therefore, that if perceptual similarity is a factor in shaping input–output mappings, its effects can be diminished or even completely masked by other factors.

A different kind of perceptual consideration that may play a role in input–output mappings is that of consonant recoverability (Browman and Goldstein, 2000; Chitoran et al., 2002; Silverman, 1997; Wright, 1996). As the perception results show, it can be difficult for English listeners to recover the presence and identity of initial stops when they are followed by either another stop or by a nasal consonant. For example, stop-nasal clusters were most confusable with C_1 deletion and C_1 change modifications. If talkers have knowledge of this difficulty they may select repair strategies to improve the recoverability of C_1. For stop-initial clusters, epenthesis is a strategy that ensures C_1 recoverability (perhaps to a greater degree than prothesis). Epenthesis was the dominant repair strategy for not just stop-initial clusters but for all cluster types. This may reflect a bias towards employing a consistent repair strategy in the production of all unattested sequences. Thus, in addition to mechanical noise in speech articulation, pressure to maintain the recoverability of stops and to have a uniform repair strategy may also influence non-native phonotactic production.

In summary, the modeling approach developed in this paper allowed us to hone in on aspects of speech production data attributable to unfaithful phonological mappings. This made it possible to evaluate an extension of the P-map hypothesis to non-native phonotactic production. The mismatch between input–output mappings and patterns of perceptual similarity fail to support for this extension of the P-map. Besides this result, our methodology revealed three facts to be explained by the grammar. First, the phonological grammar of English allows faithful input–output mappings for unattested stop-initial clusters more often than it allows faithful input–output mappings for unattested fricative-initial clusters. This fact about the relative markedness of clusters emerges only after distilling cases of epenthesis from lower level production errors. Second, for all manner combinations, epenthetic mappings are the most frequent unfaithful mappings. This result suggests a bias towards applying a uniform repair strategy across all manner combinations in the corpus. Lastly, we found that prothesis occurs more often for fricative-initial clusters than for stop-initial clusters and discussed some complications for an explanation of this result in terms of the P-map hypothesis.

6. Conclusion

Relating experimental data to phonological theory may require consideration of grammar-external influences on behavior. In this paper we developed a computational model to aid in linking speech production data to the output of the phonology. The model revealed that a large proportion of voiced vocoids surfacing between consonants are due to imprecision in consonant timing as opposed to phonological epenthesis. This finding produced a straightforward explanation for the difference in insertion percentages between stop-initial clusters and fricative-initial clusters. Under identical noise conditions, production of a fricative-initial cluster is less likely to reveal a period of inter-consonantal voicing than production of a stop-initial sequence. This is attributed to differences in the internal temporal properties of stops and fricatives.

In order to evaluate the P-map hypothesis, we used the model as an analytical tool to sort the insertion cases in the data into instances of epenthesis and instances of articulatory mistiming. We established through a discrimination experiment that the most confusable input–output pairs differed according to the manner combination of the target consonant clusters. Fricative-stop clusters, FS, were most perceptually similar to prothetic forms, əFS. Stop-stop clusters, SS, were comparably similar to both prothetic forms, əSS, and epenthetic forms, SəS. Despite this nuanced pattern of perceptual similarity, the most frequent unfaithful mapping for both fricative-initial clusters and stop-initial clusters yielded CəC sequences. The mismatch between patterns of perceptual similarity (FS most similar to əFS) and patterns of unfaithful mapping (FS mapped to FəS) is inconsistent with a potential extension of the P-map hypothesis to the production of phonotactically unattested sequences. We argued instead that a number of factors, including the recoverability of input consonants, pressure to maintain a uniform repair strategy, and motor noise in speech production, all contribute to shaping patterns in the experimental data.

More broadly, the results indicate that, in the presence of appropriate analytical tools, speech production experiments can be highly informative in evaluating phonological theory.

Acknowledgments

The authors would like to thank members of the NYU Phonetics and Experimental Phonology Lab, the MARCS Early Career Researchers group, Catherine Best, and the audience at NELS 40 at MIT for their feedback on this work. We would also like to thank Marcos Rohena-Madrazo for assistance with data collection. This research was supported by NSF CAREER grant BCS-0449560 to the second author.

Appendices A and B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.lingua.2011.03.003.

References

Baertsch, K., Davis, S., 2009. Strength relations between consonants: a syllable-based OT approach. In: Nasukawa, K., Backley, P. (Eds.), Strength Relations in Phonology. Mouton de Gruyter, New York, pp. 293–324.

Berent, I., Lennertz, T., Smolensky, P., Vaknin-Nusbaum, V., 2009. Listeners' knowledge of phonological universals: evidence from nasal clusters. Phonology 26, 75–108.

Berent, I., Steriade, D., Lennertz, T., Vaknin, V., 2007. What we know about what we have never heard: evidence from perceptual illusions. Cognition 104, 591–630.

Best, C., Tyler, M., 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, M., Bohn, O.-S. (Eds.), Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production. Johns Benjamins, Amsterdam, pp. 13–34.

Bosch, A., de Jong, K., 1997. The prosody of Barra Gaelic epenthetic vowels. Studies in the Linguistic Sciences 27, 1–15.

Broselow, E., Chen, S.-I., Wang, C., 1998. The emergence of the unmarked in second language phonology. Studies in Second Language Acquisition 20, 261–280.

Browman, C.P., Goldstein, L., 1990. Gestural specification using dynamically-defined articulatory structures. Journal of Phonetics 18 (3), 299–320.

Browman, C.P., Goldstein, L.M., 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. Les cahiers de l'lCP, Bulletin de la Communication Parlee 5, 25–34.

Catford, J.C., 1977. Fundamental Problems in Phonetics. Indiana University Press, Bloomington.

Chitoran, I., Goldstein, L.G., Byrd, D., 2002. Gestural overlap and recoverability: articulatory evidence from Georgian. In: Gussenhoven, C., Warner, N. (Eds.), Laboratory Phonology 7. Mouton de Gruyter, Berlin, New York, pp. 419–447.

Cho, T., Ladefoged, P., 1999. Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics 27 (2), 207–229.

Davidson, L., 2006a. Phonology, phonetics, or frequency: influences on the production of non-native sequences. Journal of Phonetics 34 (1), 104–137.

Davidson, L., 2006b. Phonotactics and articulatory coordination interact in phonology: evidence from non-native production. Cognitive Science 30 (5), 837–862.

Davidson, L., 2010. Phonetic bases of similarities in cross-language production: evidence from English and Catalan. Journal of Phonetics 38 (2), 272–288.

Davidson, L., Shaw, J., Sources of illusion in consonant cluster perception (pp. 1–26), submitted for publication.

Dell, F., Elmedlaoui, M., 1996. Nonsyllabic transitional vocoids in Imdlawn Tashlhiyt Berber. In: Durand, J., Laks, B. (Eds.), Current Trends in Phonology: Models and Methods. University of Salford Publications, CNRS, Paris and University of Salford, pp. 219–246.

Dell, F., Elmedlaoui, M., 2002. Syllables in Tashlhiyt Berber and in Moroccan Arabic. Kluwer Academic Publishers, Dordrecht, Netherlands, and Boston, MA.

Docherty, G., 1992. The Timing of Voicing in British English Obstruents. Foris, Berlin.

Eckman, F.R., 1977. Markedness and the contrastive analysis hypothesis. Language Learning 27, 315–330.

Flege, J.E., 1995. Second-language speech learning: theory, findings and problems. In: Strange, W. (Ed.), Speech Perception and Linguistics Experience: Issues in Cross-language research. York Press, Timonium, MD, pp. 229–273.

Fleischhacker, H., 2005. Similarity in Phonology: Evidence from Reduplication and Loan Adaptation. Los Angeles, UCLA.

Gafos, A., 2002. A grammar of gestural coordination. Natural Language and Linguistic Theory 20, 269–337.

Goldrick, M., 2011. Utilizing psychological realism to advance phonological theory. In: Goldsmith, J., Riggle, J., Yu, A. (Eds.), Handbook of Phonological Theory. 2nd edition. Blackwell.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Hall, N., 2006. Cross-linguistic patterns of vowel intrusion. Phonology 23, 387–429.

Kabak, B., Idsardi, W., 2007. Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? Language and Speech 50, 23–52.

Lado, R., 1957. Linguistics Across Cultures. The University of Michigan Press, Ann Arbor, MI.

Lin, Y., 1997. Syllabic and moraic structure in Piro. Phonology 14, 403–436.

Lisker, L., Abramson, A., 1964. A cross-language study of voicing in initial stops: acoustical measurements. Word 20, 384–422.

Lisker, L., Abramson, A., 1967. Some effects of context on voice onset time in English stops. Language and Speech 10, 1–28.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide, 2nd ed. Lawrence Erlbaum Associates, Mahwah, NJ.

McCarthy, J.J., Prince, A., 1995. Faithfulness and reduplicative identity. In: Beckman, J., Walsh Dickey, L., Urbanczyk, S. (Eds.), University of Massachusetts Occasional Papers in Linguistics 18. GLSA Publications, Amherst, MA, pp. 249–384.

Pisoni, D., 1973. Auditory and phonetic codes in the discrimination of consonants and vowels. Perception & Psychophysics 13, 253–260.

Prince, A., Smolensky, P., 2004. Optimality Theory: Constraint Interaction in Generative Grammar. Blackwell Pub., Malden, MA.

Ridouane, R., 2008. Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber. Phonology 25 (02), 321–359.

Shaw, J.A., Gafos, A., Hoole, P., Zeroual, C., 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. Phonology 26, 187–215.

Shaw, J. A., Gafos, A., Hoole, P., & Zeroual, C. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters (pp. 1–28), submitted for publication.

Shaw, J.A., Gafos, A.I., 2010. Quantitative evaluation of competing syllable parses. In: ACL SIGMORPHON-11, 11th Meeting of ACL Special Interest Group in Computational Morphology and Phonology, Uppsala, Sweden, p. 8.

Silverman, D., 1997. Phasing and Recoverability. Garland, New York.

Steriade, D., 1990. Gestures and autosegments: comments on Browman and Goldstein's paper. In: Kingston, J., Beckman, M. (Eds.), Papers in Laboratory Phonology. Cambridge University Press, Cambridge.

Steriade, D., 2008. The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. In: Hanson, K., Inkelas, S. (Eds.), The Nature of the Word: Studies in Honor of Paul Kiparsky. MIT Press, Cambridge.

Strange, W., Shafer, V.L., 2008. Speech perception in second language learners: the re-education of selective perception. In: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and Second Language Acquisition. John Benjamins, Philadelphia, pp. 153–192.

Werker, J., Logan, J.S., 1985. Cross-language evidence for three factors in speech perception. Perception & Psychophysics 37 (1), 35–44.

Westbury, J.R., 1994. X-ray Microbeam Speech Production Database User's Handbook. University of Wisconsin, Madison, WI.

Wright, R. (1996). *Consonant Clusters and Cue Preservation in Tsou*. Unpublished Ph.D. dissertation, UCLA.