Utterance preparation and Stress Clash: Planning prosodic alternations

SAM TILSEN

Abstract: Abstract: Models of speech planning and production allow for the prosodic structure of an utterance to be only partially built when the utterance is produced. Previous studies have found evidence that higher-level prosodic units exert a stronger influence in relatively prepared speech compared to unprepared speech. A related issue is whether prosodic alternations are likewise influenced by preparation. This chapter reviews phonological and phonetic treatments of a clash-driven prosodic alternation known as the "rhythm rule", and reports on a preliminary analysis of an experimental study. The experiment tested the hypothesis that the extent to which an utterance has been prepared will affect the magnitude of prominence reduction that typically occurs in the rhythm rule. Durational decreases due to clash were observed only in prepared utterances; in unprepared speech clash unexpectedly resulted in an increase in segmental durations. These results suggest that variation due to prosodic alternations is substantially influenced by the extent to which an utterance has been prepared, and hence provide a new direction for exploration of the interaction between prosody and time-constraints on planning.

1 Introduction

Phonological representations tend to promote the assumption that prosodic structure is fully built prior to the initiation of an utterance. This may be due to emphasis placed on changes that occur in deriving surface forms from underlying ones, as well as the relatively static nature of the representations themselves. Here "prosodic structure" is construed broadly to include the organization of prosodic boundaries and relations of relative prominence. Prominence is associated with individual syllables and is expressed with a combination of acoustic variables, particularly duration and fundamental frequency (F0). Phonological factors such as metrical accent (or stress), intonational pitch accent, and prosodic boundaries, all contribute to prominence.

Some utterances may be quite short and consist of only a single word, syllable, or even a lone gesture; however, for current purposes, it is desirable to consider longer utterances consisting of multiple words. There is a tendency in conversational speech for words in such utterances to be grouped into phrases, and this grouping is often associated with prosodic patterns involving metrical and/or intonational pitch accent. Importantly, when grouping of words into higher-level prosodic phrasal units occurs, there is a potential for alternations in how stress and pitch accent are realized in a given word, compared to how they are realized when the word is produced in isolation. Assuming that most phrases are not lexically stored, these alternations must be computed on-line, either prior to or during production. While it is possible for boundary organization and prominence to be fully determined prior to the production of a phrase, current models of speech production allow for variation in the extent to which prosodic structure is built. This paper will explain how two types of speech production models - incremental and dynamical models - predict variation in prosodic structure to arise from the amount of time available for utterance preparation. Specifically, the focus here is on a prosodic alternation known as the "rhythm rule." This alternation involves a difference in the prominence pattern of a word when it followed by another word with initial prominence, e.g. "se.ven.TEEN ga.ZELLES" vs. "SE.ven.teen GEC.kos". Examples 1-3 below illustrate the rhythm rule in several contexts. In examples 1a, 2a, and 3a, the primary stress in the underlined word falls on the final syllable, with secondary stress falling on a preceding syllable. Here there is no application of the rhythm rule since the initial syllable of the following word is unstressed. In contrast, examples 1b, 2b, and 3b, illustrate how primary stress occurs earlier in the word (marked in bold) when it is followed by a word with initial stress. Various phonological approaches to representing this alternation are considered in section 1.2. There are two general requirements for the alternation to occur: first, the words participating in the rhythm rule (the bracketed words in examples 1-3), must be grouped together into a prosodic phrase, and second, there must be a stress or pitch accent "clash", where juxtaposition of isolation forms of the words gives rise to a pattern of adjacent primary stresses or pitch accents within the phrase.

An experimental study was conducted to investigate whether the rhythm

rule is manifested differently in prepared than in unprepared speech. Because models of production posit that utterances can be initiated without fully determined prosodic structure, it is possible that the phonetic manifestations of the rhythm rule may differ between utterances that have been planned to a greater or lesser extent. A preliminary analysis of experimental results is presented, which shows that changes in relative syllable prominence associated with the rhythm rule are influenced by the extent to which an utterance has been prepared. In other words, whether speakers produce an utterance as soon as they are able to, or delay production for a brief period of time, influences whether prosodic alternations occur. This result suggests that future studies of prosody and articulation take into account the temporal dynamics of utterance preparation.

1.1 Prosodic structure in models of speech planning

Most models of speech planning and production allow for utterances to be produced with incompletely built prosodic structure. Both incremental and dynamical models are discussed here in order to familiarize the reader with different approaches to deriving preparation-related predictions. In incremental models, structure can be reformed iteratively during utterance planning; as more lexical information is retrieved from memory, processing modules are able to construct more detailed representations and potentially alter those representations. Utterances can be produced before all potentially relevant information has had an opportunity to influence the structural representations computed by the processors. Dynamical models of planning associate structural units with "planning systems", which are modeled by real-time continuous variables. These variables can interact with each other through coupling, and the interactions can result in changes that influence the magnitudes and durations of articulatory gestures. The process of fully activating units takes some time, and the effects of their interactions emerge over time, hence the system of dynamical variables requires some time to reach a steady state. If an utterance is produced before planning systems have fully stabilized, prosodic structure and alternations may be only partially realized.

Levelt (1989) and Levelt et al. (1999) present an incremental model of

speech planning which is quite comprehensive in that it accommodates a range of segmental and prosodic phenomena. In the model, metrical structure is retrieved at an early stage from lexical entries. This occurs in a metrical spell-out routine, which accesses lexically stored information regarding the number of syllabic peaks in a word and the location of word accent (primary stress). Metrical spell-out provides input to a prosody generator, in parallel with segmental spell-out. The prosody generator computes phonetic parameters that serve as input to phonetic spell-out; these phonetic parameters include duration, intensity, and pitch for a given syllable frame, and they are computed based upon the structural representations of metrical stress, pitch accent, and prosodic boundaries that are available to and computed by the prosody generator (Levelt, 1989, p. 367). To accommodate prominence alternations in the rhythm rule context, the prosody generator is able to "look ahead" to the metrical structure of following words and compute a change in prosodic structure. The lookahead process must also take into account prosodic phrase structure, in order to avoid triggering metrical alternations when a clash straddles a phrase boundary.

The rhythm rule is categorical but not obligatory in the Levelt (1989) model. Information regarding the metrical pattern of the upcoming word may not be available when the prosody generator computes phonetic parameters, "because of a high speech rate, or for some other reason" (Levelt, 1989, p. 374). The model thus makes the following prediction: the rhythm rule either will or will not be implemented, depending upon whether information regarding the metrical structure of the upcoming word is available when the prosody generator computes phonetic parameters for a given word and outputs them to phonetic spell-out. In either case, the phonetic outcome is dependent upon whatever partially-built metrical structure is available for input to the prosody generator when it computes its output phonetic parameters.

Dynamical planning models also allow for utterances to be produced with incompletely built prosodic structure. Although less comprehensive in scope than the incremental model of Levelt (1989), several models have extended the task-dynamic framework of articulatory phonology (Browman and Goldstein, 1992; Saltzman and Munhall, 1989) to incorporate prosodic planning dynamics. These models are based upon hierarchical networks of coupled oscillators. In these approaches prosody-

related patterns of duration and relative timing in production arise from changes in frequency-locking or strength of phase-coupling between syllables, feet, and prosodic phrases, in some cases in conjunction with the presence of special prosodic gestures that slow the production system (Byrd and Saltzman, 2003; Saltzman et al., 2008; Tilsen, 2009). The framework developed in Tilsen (2011a,b) allows for both phase-coupling and amplitude-coupling between oscillatory planning systems at various levels of the prosodic hierarchy. Amplitude-coupling forces influence the relative activations of syllable and gestural planning systems, which in turn can influence gestural magnitudes and durations. Due to interactions between amplitude-coupling forces and relative phase, these forces can promote rhythmic alternation of syllable prominence (Tilsen, 2011b). Crucially, the coupling forces take some time to exert a maximal effect. This allows for the possibility that utterances produced with a relatively short period of time for planning may not exhibit fully stabilized prosodic dynamics.

The potential for speech to be produced when prosodic planning systems have not fully stabilized is similar to incomplete building of structure; however, there is a potentially important conceptual distinction to be drawn. The dynamical notion of structure building assumes a gradient view of structure in which prosodic "units", such as stress or pitch accents, are dynamical systems that exhibit continuous variation in activation over time. This differs from the more reified metaphor of a grid in which prominence-objects are stacked in association with syllable slots and tone-units are associated to those slots. Incremental processing favors viewing structure as a composition of objects in space, since it relies fundamentally on the concept of retrieval of information from memory. The things that are retrieved are countable, e.g., a metrical pattern is a number of syllable peaks; changes in the pattern are formulated by rules that move or alter structural objects such as grid marks or pitch accents. In contrast, the dynamical conception of what is "retrieved" from memory involves a pattern of relative activation and phase, along with coupling forces; subsequent changes to the pattern ("structure building") arise from the intrinsic dynamics of systems and their coupling interactions Tilsen (see 2011a,b) for detailed illustrations of the approach). The current experiment does not aim to resolve between these competing views of speech planning, but the discussion section 4 will consider

whether the two approaches make differing predictions regarding the effects of preparation.

1.2 Experimental evidence for an influence of preparation

There are several studies which suggest that that the amount of time available for utterance preparation has an influence on the organization of prosodic structure. Numerous experiments by Sternberg et al. (1978, 1988) have shown that in prepared utterances the number of feet (or stress groups), i.e., the "length of the utterance", has a linear effect on the reaction time (RT) to initiate the utterance. They interpreted this effect to arise from a series of selection and execution processes. The selection process draws units from a motor program buffer and the execution process produces the associated motor commands. All feet in an utterance are stored as separate units in the buffer. The more units that occupy the buffer, the longer it takes to select the first unit in the utterance, and hence RT increases linearly with the number of feet in the buffer. Previously selected units remain in the buffer, so that all selection processes during an utterance are affected equally. The durations of produced units depend on the time-courses of both selection and execution processes; hence foot durations increase linearly with utterance length, because the selection of each foot will take longer if there are more of them. As a direct consequence of this, utterance length has a nonlinear effect on the duration of the utterance as a whole: the effect of utterance length on the selection of each foot is multiplied by the length of the entire utterance, i.e., the number of feet in the utterance.
More recently, similar effects have been observed in association with the number of prosodic words in an utterance. Wheeldon and Lahiri (1997) conducted a study in Dutch that suggests preparation time plays a key role in influencing the building of prosodic structure. Their study examined the effect of the number of prosodic words in an utterance on the RT to initiate the utterance. They compared utterances with two and three prosodic words in which the number of syllables was held constant. Analogous example utterances in English are "I seek the water" (two prosodic words) and "I seek fresh water" (three prosodic words).

The contrast derives from the propensity for the determiner to cliticize to the prosodic word associated with the preceding verb, whereas the adjective constitutes its own prosodic word. What makes this study particularly relevant is that Wheeldon and Lahiri conducted two versions of their experiment: one in which speakers had plenty of time to prepare the response, and another in which preparation time was substantially restricted. In the fully prepared response experiment, reaction times patterned in accordance with the Sternberg findings: RTs were greater for three-word utterances than two-word utterances. However, in the restricted preparation experiment, reaction times were not affected by utterance length. Instead, reaction times increased with the complexity of the initial prosodic word (i.e., the number of syllables in the word): the two-word utterances took longer to initiate than the three-word utterances, because the first prosodic word contained an additional syllable, the cliticized determiner. The absence of the prosodic word effect in the unprepared design suggests that higher-level phonological phrase structure requires more planning time to be fully assembled, and shows that prosodic structure need not be completely built prior to utterance initiation.

Regarding effects of utterance preparation on the segmental level of the prosodic hierarchy, it has been shown that articulatory posturing for the initial gesture(s) of a word occurs well in advance of acoustic response onset in delayed-cue naming tasks. For example, Kawamoto et al. (2008) showed that the extent of articulatory preparation increases with the amount of time between the stimulus and response cue in delayed naming. Even explicit instructions not to initiate articulation prior to the cue did not eliminate this effect. Hence the articulation of response-initial gestures can be substantially influenced by preparation, and this reinforces the notion that temporal differences in planning can manifest in articulatory variation.

The effect of utterance preparation also emerges in comparisons between choice-RT and simple-RT paradigms. In choice-RT paradigms, the response action is not known until the imperative "go" stimulus occurs, typically because the imperative stimulus also informs the subject of what response action they will perform. Hence the response cannot be prepared prior to the go-cue. In simple-RT paradigms the subject is informed of what action they will perform prior to the go-cue, as was the

case in the Sternberg et al. (1978, 1988) experiments. Klapp (1995, 2003) demonstrated that the effect of the number of units on response latency is only present in simple-RT paradigms, not in choice-RT paradigms. This pattern held true for a variety of response types: button presses, alphabet letters, words, and pseudowords. Utterance preparation is related to the choice-/simple-RT distinction, since the basic difference between the two conditions is the amount of time that a response is planned prior to its execution.

1.3 Phonological representations of prominence and the rhythm rule

Of primary concern here is the phonological representation of prosodic patterns associated with the rhythm rule (also: "stress shift"), which is driven by stress clash. Clash-related phenomena of this sort are interesting to investigate because they derive from on-line building of prosodic structure: clash does not generally arise from lexical memory, since most multi-word phrases are not lexically stored. Hence the phonetic manifestations of clash can serve as an index for the influence of preparation on prosodic structure building and on-line implementation of prosodic alternations. The rhythm rule pattern involves the early location of prominence in words where a secondary stress precedes a primary stress, such as "Cornell", "Japanese", and "Minnesota". When these words are produced phrase-finally or in isolation, their most prominent syllables are the ones with primary stress, as expected. However, when they are followed in a phrase by a word with primary stress on the initial syllable, the location of the most prominent syllable appears to "shift" earlier, to the preceding syllable that had secondary stress. Similar patterns have been observed in a number of languages, including Polish (Hayes and Puppel, 1985), Italian, Greek, Catalan (Nespor and Vogel, 1986), Biblical Hebrew (Churchyard, 1999), Bedouin Hijazi Arabic (Al-Mozainy et al., 1985), and Afghan Persian (Bing, 1980).

There are two main approaches to the phonological representation of prominence (Shattuck Hufnagel et al., 1994; Gussenhoven, 2011) and within these, several alternative accounts of how accentual variation arises. One approach, which Gussenhoven (2011) refers to as the "infi-

nite stress view" is based on the idea that accentual differences between words are the same type of phenomenon as accentual differences within words. This conception of prominence is unitary, in that no clear distinction is drawn between the contributions of word-stress and intonational pitch accent. In this view prominence values are associated with each syllable, and rules specify changes to those values - this was the approach employed by Chomsky and Halle (1968). The idea that all prominence is similar favors the use of a single representational formalism for within-word and across-word prominence. One way to express this is with metrical grid marks (Liberman and Prince, 1977; Selkirk, 1984; Nespor and Vogel, 1986), shown in Figure 1a, where the heights of grid marks express their relative prominence. A different approach is to

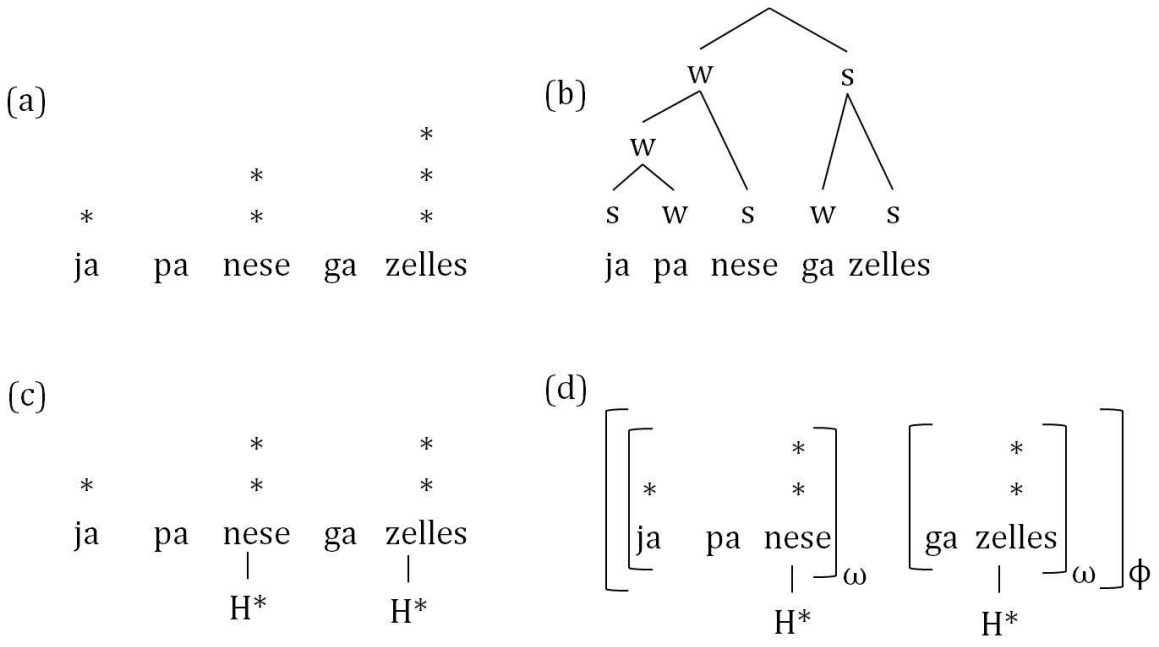

Figure 1: Phonological representations of prominence. (a) metrical grid; (b) metrical tree; (c) metrical grid with pitch accent tier; (d) bracketed grid with pitch accent tier.

use hierarchical metrical tree structures (Liberman, 1975; Liberman and Prince, 1977), which through strong or weak labeling of binary branching nodes can represent both relative prominence and grouping relations, shown in Figure 1b. Both metrical trees and grids are examples of the infinite stress view because they employ a common representational mechanism for within-word and across-word prominence: both frameworks conflate pitch accent and metrical stress in representing promi-

nence.

An alternative representational approach - the "pitch accent view" (Gussenhoven, 2011) - is based on the idea that intonational pitch accents are phonologically autonomous from word stress (Bolinger, 1958; Vanderslice and Ladefoged, 1972; Horne, 1990; Gussenhoven, 1991; Shattuck Hufnagel et al., 1994). Bolinger (1958) proposed that there are two pitch accents in a phrase: a "themic accent" occurring early in the phrase (a pre-nuclear accent), and a "rhemic accent" occurring late in the phrase, which can be equated with the main, or nuclear accent. In the pitch accent view, intonational pitch accents can be associated with syllables having unreduced vowels and can thereby imbue them with additional prominence. Pitch accents thus coexist with primary and secondary metrical stress levels, as shown in Figure 1c. The choice of which pitch accent to associate to a syllable is an independent issue. In English, there are a number of options (H*, H*L, L*H, etc.), only one is shown here.

There are a variety of hybrid approaches to representing metrical and prosodic structure, and a full review is not possible here. The reader should note that some authors use upper-level tiers in a grid to represent positions to which pre-nuclear pitch accents and nuclear pitch accents are associated (Beckman and Edwards, 1994), with the assumption that these upper-level marks do not contribute prominence in the same way as the lower-level marks. It is furthermore common to incorporate labels into a tree structure or brackets into a grid (Hayes, 1995; Nespor and Vogel, 1986), in order to represent the locations of prosodic word ω and phonological phrase boundaries ϕ, as shown in Figure 1d.

Many phonological treatments of the rhythm rule are based upon notions of stress clash and subsequent repair. Adjacent prominent syllables create a clash, for which some form of reparation is desirable. In metrical grids based on a unitary conception of prominence, the repair involves a transfer of a grid mark to a preceding syllable (Liberman and Prince, 1977), as shown in Figure 2a. In metrical trees, the shift has been conceptualized as the exchange of a strong and weak node (Kiparsky, 1979), shown in Figure 2b. There is also a more inclusive construal which allows for non-adjacent primary stressed syllables to clash when no secondary stress intervenes between them, as in "Minnesota apartments" (clash) vs. "Minnesota abbreviations" (no-clash). In pitch-accent views, the repair can be analyzed as a deletion of an accent or early occurrence

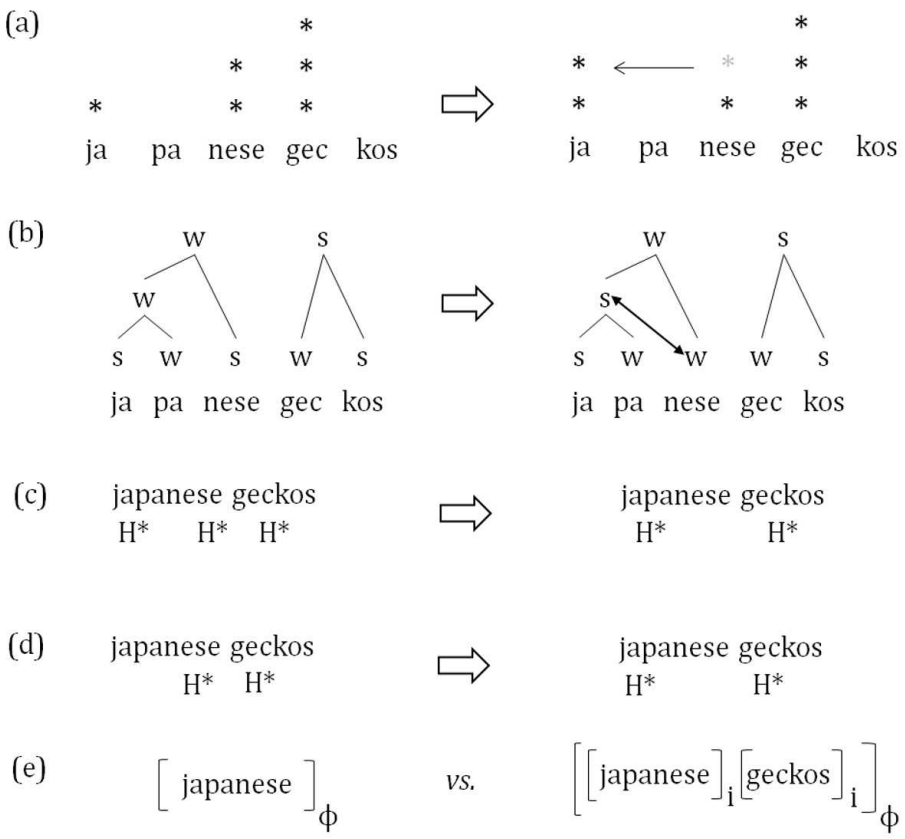

Figure 2: Phonological representations of the rhythm rule. (a) prominence-transfer in a metrical grid; (b) strong-weak label transfer in a metrical tree; (c) pitch accent deletion; (d) early pitch accent; (e) boundary insertion.

of an accent. Gussenhoven (1991, 2011) provides an accent deletion account, shown in Figure 2c. In this analysis both the primary and secondary stressed syllables have pitch accents, and a "rhythm-induced accent deletion" rule deletes medial accents in the phonological phrase. Note that this rule can apply to any phrase-medial accent, regardless of whether there is a clash induced by adjacency to a following accent. Bolinger (1965) presents an early accent account, in which a prominence occurs early due to a desire to have a prenuclear pitch accent early on in the phrase. This account can also be construed to involve a movement of the pitch accent, analogous to the movement of metrical stress.

Early pitch accent and accent deletion analyses are not based on a repair of metrical rhythm. The early location or deletion of pitch accent does not serve to avoid metrical stress clashes between grid marks. However, these patterns can be conceptualized as a repair of the "rhythm of pitch accent", which Bolinger (1965) distinguished from metrical rhythm. Im-

portantly, the pitch accent approaches do not strongly differentiate between clashing and non-clashing stress contexts: they predict that the contribution of pitch-accent to prominence in "Japanese geckos" will be the same as that of "Japanese gazelles" and "Japanese tyrannosaurus".
Beckman and Edwards (1994) describe several accounts of rhythm rule prominence patterns, one that is consistent with deletion, one consistent with early accent, and a third based on prosodic boundaries. In one account, they suggest that the first accent in a word with two prenuclear accents is perceived as stronger because there is an utterance-initial rise in pitch to the first accent. This is consistent in a perceptual sense with deletion of a medial accent, but does not necessarily involve a phonological deletion per se. In the second account, they suggest that a pitch accent is associated with the first accentable syllable in the phrase due to an impetus to have pre-nuclear accent occur as early as possible before the nuclear accent. This account is concordant with the early accent account of Bolinger (1965, 1985).
Phrase boundary structure is crucial in all of the pitch accent accounts. Both deletion of pitch accent and early location of pitch accent require an absence of preceding material in the phrase. Medial deletion only applies in non-phrase-initial words, and early accent should occur in the first word in a phrase. Hence if a word is not phrase-initial, then no pitch accent on a secondary stressed syllable is expected under either account.
A third proposal of Beckman and Edwards (1994) is that apparent early prominence occurs when the nuclear accent associated with a phrase-final boundary is no longer present, as shown in Figure 2e. This account differs somewhat from the preceding ones in that it depends on the presence or absence of a nuclear accent in the word with variable accent.
The accounts depicted in Figure 2a-e are similar in that they implicate phonological structure or the presence of stress or pitch accent. However, they differ in two regards. First, the unitary prominence accounts (2a,b) describe repairs that serve metrical (stress-based) rhythm. In contrast the pitch accent accounts describe repairs that result in greater separation of pitch-accents (2c, d), or in the case of (2e), patterns of accentuation arising from the organization of prosodic boundaries. Second, the repairs in the unitary prominence accounts are conditioned by specific clash configurations (primary stresses adjacent or separated by one syllable); in contrast, the pitch accent views hold that clash of

primary stresses is not responsible for the early occurrence of prominence. In addition to accounts based solely upon metrical stress or intonational pitch accents, there have been a number of integrated accounts which rely upon patterns in both word-level and phrasal/intonational domains. Some possibilities along these lines are discussed by Shattuck Hufnagel et al. (1994). For example, it could be the case that the deletion or movement of a pitch accent is dependent upon a deletion or movement of stress in the metrical grid. Another possibility implies the opposite direction of causation: changes in locations of pitch accents trigger changes in within-word metrical patterns.

1.4 Phonetic studies of the rhythm rule

A key question in phonetic investigations of the rhythm rule is what the phonetic correlates of prominence are. In the unitary view of prominence, variables such as duration, intensity, and F0 are all correlates of prominence; there is no dissociation of F0 with other variables, because prominence is a monolithic phenomenon. In contrast, a number of researchers have argued for drawing a distinction between direct correlates of word stress, pitch accents, and prosodic boundaries. In these views, the correlates of stress are generally considered to be increased duration, vowel quality, and intensity, whereas F0 patterns arise most directly from pitch accents and tones associated with prosodic boundaries. This distinction is accepted by a number of researchers (Beckman and Edwards, 1994; Gussenhoven, 1991; Shattuck Hufnagel et al., 1994). Beckman and Edwards (1994) point out that many studies of stress have conflated pitch accent and prosodic boundary structure with metrical stress. For example, they attribute the association of F0 with stress to a misunderstanding of experimental research in Fry (1958), which showed F0 as the primary cue for distinguishing stress pairs such as per.MIT and PER.mit. The F0 patterns were due to association of nuclear pitch accent when these words are produced or heard phrase-finally.

A complicating factor in phonetic analyses of prominence is that pitch accents, phrasal accents, and prosodic boundaries have effects on duration independent of metrical stress. For example, prosodic boundaries are known to be associated with longer durations of segments and articulatory gestures (Klatt, 1975; Nakatani et al., 1981; Beckman and Ed-

wards, 1994; Byrd and Saltzman, 1998, 2003; Byrd, 2000). The extent to which this additional duration is a "direct" consequence of the boundary, or an indirect consequence of the presence of a nuclear pitch accent, phrasal accent, or boundary tone, is not entirely clear. The opposite direction of interaction - stress influencing F0 independently of intonation - has not been as strongly supported in the literature, although it should not be ruled out given that the presence of stress on a syllable is typically considered a precondition for associating pitch accent.

Due to the above factors, durational patterns cannot be directly attributed either to metrical structure or intonational structure. This results in an unavoidable multiplicity of hypotheses when the potential structural source(s) of experimental effects are unknown. More specifically, if the occurrence of intonational accents, phrasal structure, and metrical stress patterns can be experimentally controlled independently, then effects on duration or F0 would be clearly interpretable. However in the more general condition where accentual patterns, phrasal structure, and metrical structure are unknown, durational effects can be interpreted as consequences of either one or several of these factors. To determine whether durational patterns are solely attributable to just one factor is only possible when independent characterizations of pitch accentuation, phrasal structure, and stress are available. Moreover, if we allow for these factors to be represented gradiently, the difficulties are exacerbated.

It is worth mentioning here that another possible repair in the rhythm rule context is an elongation of the first stressed syllable in the clashing pair. Liberman (1975) suggested this as a means of promoting rhythmic spacing of stresses. By lengthening this syllable, its peak would be further away from the subsequent stress on the following word. It is also possible to interpret this repair as a consequence of insertion of an additional silent beat between words (Selkirk, 1984). Durational lengthening is an interesting possibility in light of the results of the current study, and notably it runs contrary to what prior investigations have found: shortening of the first primary stressed syllable in a clashing pair.

Limited conclusions can be drawn from phonetic studies of the rhythm rule due to two types of confounds detailed below. The available evidence suggests that metrical stress clash results in a decrease in prominence of the primary stressed syllable (σ_p) in a bipedal word, as opposed to a transfer of prominence to a preceding syllable with secondary stress

(σ_s). Hence in a clashing context, σ_p in words like "Cornell", "Japanese", and "Minnesota" experiences a decrease in duration and F0, while prominence on σ_s remains relatively unchanged. For reasons described above, the decrease in prominence may have several explanations: (1) decrease in the metrical stress level of the word, (2) deletion or early placement of a pitch accent, possibly with an indirect influence on duration, (3) change in prosodic boundary structure, or (4) some combination of the preceding. In the following, we will refer to a bipedal word in a rhythm rule context as the "target word". Table 1 summarizes previous research

Table 1: Phonetic studies of stress clash: σ_s = syllable with secondary stress, σ_p = syllable with primary stress, studies are abbreviated as follows: CE86=Cooper and Eady (1986), H90=Horne (1990), GW95=Grabe and Warren (1995), VBH95=Vogel et al. (1995).

	σ_s	σ_p-σ_s	σ_p/σ_s	σ_p
duration		↓ GW95		↓ GW95, VBH95, H90
F0 peak				↓ CE86, H90
F0 average		↓ GW95		↓ VBH95
F0 range		↓ GW95		↓ GW95
Intensity	(↓ GW95)		↓ GW95	↓ GW95

on the effects of stress clash, distinguishing between effects on phonetic variables associated with the primary and secondary stressed syllables in the first word of the clashing pair. The clashing and non-clashing contexts contrasted in these studies differ in important ways.

Grabe and Warren (1995) used sentence pairs in which the location of a phrasal boundary was manipulated, e.g., the sentence "when my father watches TV soaps, they are his favorite" contains a clash between "TV" and "soaps", but the sentence "when my father watches TV, soaps are his favorite" contains no such clash due to an intervening phrasal boundary. Horne (1990) contrasted isolation productions of words such as "Dundee" to productions in a clashing context, such as "Dundee tartan". Vogel et al. (1995) contrasted clashing/non-clashing contexts by varying the location of stress in the second word in the clashing pair, e.g. "Japanese clients" (clash) vs. "Japanese canoes" (no-clash). These word

pairs were produced as subject NPs in a sentence and were preceded by one of three contexts: no preceding word form, a preceding determiner, or a preceding NP modifier, e.g. "Jack's Japanese clients". Cooper and Eady (1986) elicited similarly clashing and non-clashing word pairs in isolation and in carrier sentences. Note that these studies used read-text response elicitation paradigms, which can be classified as prepared speech.

All of the studies found evidence that clash decreased prominence in σ_p. Grabe and Warren (1995); Vogel et al. (1995); Horne (1990) observed decreased duration in σ_p, while Cooper and Eady (1986) did not measure duration in this syllable. Note that Vogel et al. (1995) measured rhyme durations, rather than syllable durations. All of the studies also observed a decrease of F0 in σ_p, albeit using different measures (peak, average, or range). Grabe and Warren (1995) furthermore observed a decrease in intensity in σ_p.

However, Grabe and Warren (1995) and Horne (1990) conflated nuclear pitch accent, boundary structure, and stress, and hence their results cannot be construed as evidence for metrical clash-induced reduction of prominence. Grabe and Warren (1995) compared clash-context productions with phrase-final productions; the latter context is expected to have nuclear pitch accent on σ_p. Hence differences in F0, intensity, and duration observed in the comparison could be due to nuclear pitch accent, the presence of a phrase boundary, and/or metrical stress alternation. Assuming that isolation productions are also phrase-final, the same confound applies to Horne (1990). In contrast, Vogel et al. (1995) and Cooper and Eady (1986) did not confound these factors; the phrasal structures they compared should be identical in the local vicinity of σ_p.

The only evidence for a metrical clash-induced increase of prominence in σ_s, which would be due to transfer of stress, was reported by Grabe and Warren (1995). They found an increase in intensity as measured by the amplitude integral over the syllable, and they furthermore observed changes in the relative durations, pitch range, and average pitch between σ_s and σ_p. These relative changes, in conjunction with a lack of absolute changes in duration/F0 in σ_s, suggests that early prominence could result from perceived changes in relative prominence, rather than a movement or augmentation of prominence. A problematic factor here is that early pitch accent could have occurred on σ_p and obscured changes

otherwise attributable to metrical clash. An early prenuclear accent of this sort would occur in both metrically clashing and non-clashing contexts, and hence may interfere with observation of effects related to metrical stress. In Vogel et al. (1995), where preceding context was systematically varied, it may have been possible to resolve this confound by looking for an interaction effect between clash and context, but the authors did not report this analysis. A corpus study conducted by Shattuck Hufnagel et al. (1994) attempted to distinguish changes in pitch accent from changes in metrical stress; the study found evidence that early placement of pitch accent occurs without corresponding durational changes associated with metrical clash. The study furthermore observed the occurrence of doubly pitch-accented words, with accents on both the secondary and primary stressed syllables of a word. These findings argue strongly for a dissociation of metrical stress and pitch accent with regard to the rhythm rule context. Shattuck Hufnagel et al. (1994) also found evidence for avoidance of clash between pitch accents, which was independent of metrical stress clash: pitch accents exhibited a tendency for placement earlier in a word when placement on the primary stress would result in adjacent pitch accents. However, the dataset available to the study was not large enough to draw conclusions about the effects of metrical stress clash independent of pitch accent.

In sum, the evidence available from phonetic studies of metrical stress clash is rather limited. Only one study, Vogel et al. (1995), has demonstrated (without apparent intonational or structural confounds) that metrical stress clash results in decrease of duration in σ_p, and two studies, Vogel et al. (1995) and Cooper and Eady (1986), have found evidence for decrease of F0. As for whether there is a transfer of stress to σ_s, it is not clear that any of the studies offer interpretable evidence either way: the designs in Grabe and Warren (1995); Horne (1990) and Cooper and Eady (1986) all potentially suffer from obscuring effects of early pitch accent because their target words were phrase-initial; Vogel et al. (1995) did not present the relevant interaction analysis that could resolve the issue.

1.5 Hypotheses

Prominence reduction in stress clash is a good candidate for testing the effects of preparation on prosodic planning, because the reduction does

not arise from lexical memory, but rather, must arise from on-line changes in planning prosodic structure. Since previous studies have shown that preparation time is required for prosodic word structure to have an influence on reaction time to initiate an utterance (Wheeldon and Lahiri, 1997), it is possible that a prosodic alternation such as clash-induced reduction of prominence may likewise depend on preparation. However, because there are several factors that influence prominence - i.e., metrical stress within words, intonational pitch accent, and prosodic boundary structure - related predictions can be generated from a multiplicity of hypotheses. Further complicating the issue, the phonological representations discussed above may involve a number of intermediate representational stages, whose characteristics may influence phonetic observables in unprepared speech. To some extent these hypotheses cannot be differentiated with the data and analyses presented here, but the results do confirm predictions of incremental and dynamical planning models that utterances can be initiated before prosodic alternations have occurred.

Because we are interested in the effect of planning time (preparation) on prosodic alternations, and because there is some evidence from previous phonetic studies that metrical stress clash reduces prominence in the rhythm rule context, our primary interest relates to the interaction of preparation and clash. In other words, the primary hypotheses involve interaction effects: how does preparation affect the phonological and phonetic consequences of clash in the rhythm rule context? To address this question, the experimental design elicited productions of target words in the middle of formulaic three-word noun phrases with or without metrical clash, such as "eleven Japanese geckos" (clashing) or "eleven Japanese gazelles" (non-clashing); furthermore, there were two experimentally controlled preparation conditions, one which encouraged preparation and another that restricted preparation (section 2 details the methodology).

The hypotheses and predictions presented below will remain neutral with regard to conceptualization of phonological representations; hence "reduction" may refer to a categorical deletion of a representational object such as a tone or a grid mark, which accords with conventional phonological representations of structure and the incremental model of Levelt (1989); or alternatively, "reduction" may refer to a gradient

diminution of the influence of units on production, which accords with dynamical models of speech planning in which phonological units are associated with activation variables.
Due to the preliminary status of the experimental analysis, pitch accent analyses have not been conducted. Hence the hypotheses must be equivocal regarding whether durational effects are due to metrical or pitch-accentual clash. The clash context potentially induces both varieties of clash: there is an adjacency between a potential pre-nuclear pitch accent on σ_p and the nuclear pitch accent on the primary stress of the following word; there is also an adjacency between the primary metrical stresses of the target word and the following word. Since the target words were embedded in three-word noun phrases, their phrase-mediality should discourage early accent placement. However, if prosodic structure is not fully formed, as might occur in unprepared speech, then each content word may constitute a separate phonological or intonational phrase; this would allow for early pitch accent placement. Such early accent placement would be a confounding factor, particularly for hypothesis 2, which regards changes in σ_s. Without a reliable intonational analysis, it cannot be determined whether an early accent occurred. Nonetheless, measurements of segmental duration and vowel intensity can still offer revealing effects related to clash and preparation. Future analyses of intonation (e.g. in the ToBI framework, Beckman et al. (2005)) will offer further insight into the experimentally observed patterns.

Hypothesis 0 (null hypothesis)
Preparation will have no effect on the phonetic manifestations of prominence. This should be the case if prosodic structure, including potential alternations in stress/pitch accent, is fully built prior to production of an utterance.

Prediction
Clash-induced decreases of duration will be the same in prepared and unprepared speech.

Hypothesis 1
Clash-induced reduction of prominence in σ_p will be heightened in pre-

pared speech compared to unprepared speech, because preparation allows for prosodic structure to be more fully built, resulting in a greater likelihood/magnitude of reduction of stress/pitch accent.

Prediction
The clash-induced decrease in duration of σ_p will be relatively larger in prepared speech compared to unprepared speech.

Hypothesis 2
Clash-induced increase of prominence in σ_s (the secondary stressed syllable in the target word) will be heightened in prepared speech, due to movement of metrical stress and/or early location of pitch accent. These effects should occur in addition to those described by Hypothesis 1.

Prediction
In addition to predictions from Hypothesis 1, there will be a clash-induced increase in duration in σ_s, and this increase will be relatively larger in prepared speech compared to unprepared speech.

The above hypotheses reflect only a subset of possible hypotheses that could be derived from the phonological interpretations of the rhythm rule discussed in section 1.3. They have been selected to represent the most plausible and distinct accounts of the phenomenon. It should be emphasized that hypotheses 1 and 2 predict an interaction between the effect of clash and preparation: the effects of clash will be greater in prepared speech than in unprepared speech. The remainder of this paper presents the experimental method used to test these hypotheses (section 2), reports the results of the experiment (section 3), and discusses interpretations of the findings (section 4).

2 Methodology

This section describes the experimental design, stimuli, tasks, and data analysis methods. Previous studies of the rhythm rule have typically elicited a variety of lexical items from speakers, treating the identity of the response word as a random factor or source of noise that is aver-

aged out in the analysis. In contrast, this study focused on just a single word: "Japanese", which is known to exhibit reduction of prominence in a clashing context. Although the focus on a single word sacrifices some generalizability, it has the advantage that many repetitions of the same word can be obtained; this results in an increase in statistical power and hence enables smaller effects to be detected.

2.1 Experimental design

Five native speakers of American English with normal speech and hearing participated in the experiment (4 females, 1 male). Each subject participated in one 1h session, during which they produced a total of 360 utterances. Each session was divided into 10 blocks of 36 trials. Utterances were recorded with a head-mounted microphone. Matlab was used for stimuli presentation and audio recording.
On each trial, three visual stimuli of rectangular shape (each approximately 40% of screen dimensions) were displayed simultaneously on a screen with a black background. Each stimulus belonged to one of three classes: a numeral, a flag (representing a nationality), or an animal. The stimuli were positioned in a triangular arrangement, with the numeral centered on the screen in the upper row and in the bottom row, the nationality and animal positioned on the left and right, respectively. There was always exactly one stimulus from each class. Furthermore, each stimulus class consisted of only a small set of images. The numeral set contained images of the numerals twenty, eleven, and seventeen. The nationalities set contained images of the national flags of Mexico, Jamaica, and Japan. The animals set contained images of a gecko, a buffalo, a gazelle, and a piranha. The images were obtained from a Google image search. For the animals, images with minimally salient backgrounds were selected. All combinations of numerals, nationalities, and animals (3 x 3 x 4) were displayed in random order in each block. Subjects responded to the visual stimuli by producing a formulaic phrase that combined the numeral, nationality, and animal into a noun phrase, e.g. "twenty Japanese geckos" or "seventeen Mexican gazelles".

Prior to the experiment, subjects were shown the flags and animals and asked to identify them. Some subjects did not recognize the Jamaican flag and were informed that it referred to the nationality "Jamaican". Several subjects also did not recognize the image of the piranha as such, and were informed of its referent. No subjects reported that any of the nationalities or animals were unfamiliar to them. Subjects were subsequently shown all of the pictures in random order and asked to name them, until they were able to name all of the pictures with 100% accuracy twice in a row. Although there is both within- and across-subject variation in familiarity with these words, the repeated presentation of the stimuli before and during the experiment is likely to minimize the effects of prior differences in familiarity.

The independent variable of preparation was controlled by manipulating the task instructions. Subjects were told before the experiment that there were two different tasks: a quick response task, and a prepared response task. For the quick response task, they were told to respond with the correct phrase as soon as they could, right after the images appeared on the screen. For the prepared response task, they were told to rehearse the response exactly once, without producing any sound or moving their mouths, and then to say it out loud. At the start of each block, they were told which instruction to follow. The immediate (relatively unprepared) and rehearsed (relatively prepared) response tasks alternated between blocks. The independent variable of clash corresponded to the presence/absence of word-initial stress on the animal noun. Hence the clash condition corresponds to utterances with "gecko" or "buffalo" and the no-clash condition to utterances with "gazelle" or "piranha".

To encourage subjects to conform to the task instructions, feedback was given when they initiated the response too early or too late. Furthermore, to promote consistency in speech rate within and across subjects, feedback was given when the duration of the response was abnormally short or long. After each trial, the responses were analyzed to estimate when the onset and offset occurred. Based on pilot work, it was determined that responses initiated before 400ms or after 1100ms in the quick response task were abnormally early/late. For the prepared response task, these limits were 1200ms and 2500ms. If the response fell outside of these limits, the subject was notified after the trial. Pilot work also showed that the range of response duration is within 900-1500ms. If the

response duration fell outside of this range, subjects were informed that their response was too quick or too slow.

2.2 Data analysis

The analysis of experimental data presented herein focuses on segmental durations and vowel intensities in the word "Japanese". This word is typically transcribed phonetically as [dʒæpənɪz]. The intensity measurements were taken separately over the vowels in the initial and final syllables of "Japanese". Intensity was measured by analysis of the vocalic energy envelope (Tilsen and Johnson, 2008), which is a 10Hz low-pass filter of the magnitude of a passband-filtered speech waveform (5^{th} order Butterworth, 400-2400Hz). The vocalic energy amplitude envelope captures slow fluctuations in signal energy that arise primarily from vocalic resonance during voiced speech. Amplitude envelopes were normalized across the production within each trial. Two types of intensity measurements were obtained from each vowel: a peak value and an average value. These values and their ratios between the first and last vowels in "Japanese" served as dependent variables in the regression analyses presented below.

To measure segmental durations, the acoustic waveform from trials with productions of "Japanese" was low-pass filtered to allow for the approximation of syllable and word boundaries, which were then checked for accuracy by visual and auditory inspection and corrected where necessary. In order to automatically locate segmental boundaries within the word "Japanese", a custom forced alignment algorithm was implemented in Matlab. For each subject, the segments in the word were hand-labeled for two trials in each clash and preparation condition. From these, an acoustic template was created for each segment by averaging Mel frequency-scaled cepstral coefficient vectors. These vectors were then cross-correlated with cepstral coefficient matrices (13 cepstral coefficients, 20ms frames, 5ms steps, 80-5000Hz) computed from the acoustic waveform of "Japanese" from all trials. Segmental boundaries were estimated as crossing-points in the cross-correlation functions of segments expected to be adjacent. Subsequently all trials were auditorily and visually inspected for segmentation accuracy with waveforms and spectrograms; in a few instances the segmentations were corrected man-

ually, but more than 99% of them were deemed accurate without correction.

Some trials were excluded from the analysis, for various reasons. The first ten trials of the first block in each session were not analyzed. Responses that occurred early or late (see section 2.1) were excluded (2.2%), as were responses that were abnormally fast or slow, i.e., shorter than 900ms (1.8%) or longer than 1500ms (5.9%). During auditory-visual inspection, responses identified as containing incorrect or disfluent productions were excluded. Most of these involved a hesitation prior to the final word in the utterance or a production of an incorrect (unelicited) word. As one might expect, errors occurred predominantly in the unprepared condition, although they did not occur frequently enough to discern any word-specific biases. With the remaining data, trials were excluded in which a segmental duration was more than 2.2 standard deviations from the mean, calculated within subjects.

Segmental durations were log transformed and analyzed using mixed-effects linear regression. The word boundary-adjacent segments of "Japanese" were excluded from this analysis because the forced alignment estimation of the word-initial and word-final boundaries was strongly influenced by the preceding/following context, due to coarticulation with segments from flanking words. The random factors in the model were SUBJECT and BLOCK nested within SUBJECT. BLOCK effects were treated as random nested effects because they generally reflect subject-specific changes in attention and practice. The fixed effects of primary interest are preparation (PREP, manipulated by task instruction) and CLASH (the presence of initial stress on the following word). Specifically, it is the magnitude and direction of the CLASH x PREP interaction that is of relevance to testing the hypotheses. Additional fixed effects included in the model were the identity of the preceding numeral and the number of syllables in the word following "Japanese". An initial regression with all interactions showed that only the CLASH x PREP interaction term significantly improved the model fit, and hence all other interaction terms were excluded from subsequent regressions. To display the effects graphically in Figure 3, durations were z-score normalized within subjects and pooled by condition.

3 Results

The main result of the experiment is a significant CLASH x PREP interaction effect observed on durations associated with segments in the primary stressed syllable in "Japanese". CLASH decreased these segmental durations in prepared speech but not in unprepared speech, which supports hypothesis 1. Unexpectedly, CLASH actually increased the durations of the same segments in unprepared speech. CLASH x PREP interaction effects were not observed in intensity measures. Section 3.1 reports the main effects of the fixed factors, section 3.2 reports the CLASH x PREP interaction effects.

3.1 Main regression effects

PREP had a significant main effect (χ^2=9.65, p<0.001) on the duration of [i], such that [i] was longer in prepared speech. No similar effects were observed on other segments. There were no significant effects of PREP on peak or average intensity measures or their ratios. One possible explanation for the restriction of the effect to [i] duration is that the relatively long duration of [i] (due to its status as the nucleus of the primary stress) endows it with the greatest susceptibility to subtle changes in speech rate associated with preparation. Consistent with this is the finding that the duration of the entire word was likewise increased by PREP (χ^2=6.46, p<0.01).

CLASH had no significant main effect on segmental durations. Regarding the vowel [i] of the primary stressed syllable, CLASH had a quite marginal effect to increase average intensity (χ^2=2.14, p=0.14). Regarding the vowel [ae] in the secondary stressed syllable, CLASH had no effect on average intensity (χ^2=0.01, p=0.92). Similar patterns held for peak intensities, but not the ratios of intensities. Although CLASH had very little effect on segmental durations overall, we see below in section 3.2 that the interaction between CLASH and PREP is responsible for this.

The identity of the numeral had significant main effects on [ae] and [n] durations, as well a marginal effect on [p]. This effect was to increase segmental durations after the numeral "eleven". This finding is not directly relevant to the hypotheses, but including the identity of the preceding word in the regression model facilitates analysis of the

effects of interest. The number of syllables in the animal noun following "Japanese" had a strong effect on the duration of [i] (χ^2=33.5, p<0.001), but not on other segments. The trisyllabic animal words were associated with decreased duration of [i]. A reasonable interpretation of this effect can be derived from the assumption that speakers exert some global control on the duration of the entire phrase. When the final word is longer (i.e., trisyllabic as opposed to disyllabic), speakers shorten the medial word in order to approximate a target phrase duration. If [i] is assumed to have a relatively adjustable duration because of its status as nucleus of a syllable with primary stress, then it would be expected to show this effect most strongly. In contrast, the regressions did not show any main (or interaction) effects on the duration of schwa, perhaps because this segment is typically the shortest and the statistical power provided by the experiment was not sufficient to detect durational variation.

3.2 CLASH x PREP interaction effects

The results support hypothesis 1, which predicted that clash would cause a greater durational reduction in the primary stress syllable in prepared speech compared to unprepared speech. This is evidenced by the presence of a significant CLASH x PREP interaction effect on the duration of [n], which constitutes the onset of the primary stressed syllable (χ^2=6.90, p=0.01). The effect on [i] alone was not significant (χ^2=1.18, p=0.28), but it was significant on the combined duration of the onset and vowel [ni] (χ^2=5.51, p=0.02). Figure 3 shows the mean z-score for each segment for the four combinations of CLASH and PREP, along with 95% confidence intervals. It can be seen that the duration of [n] is significantly lower in a prepared speech clashing context than in a prepared speech non-clashing context. The absolute magnitude of this effect for individual speakers is on average approximately 15ms. However, a larger, opposite effect was unexpectedly observed in unprepared speech: [n] duration was lengthened by CLASH, where the average effect size across speakers was around 35ms. The same pattern holds marginally in [i]. The source of this pattern is further discussed in section 4. The results also provide relatively weak support for hypothesis 2. This hypothesis is derived from transfer-based phonological analyses of the rhythm rule, which predict that CLASH would result in heightened prominence

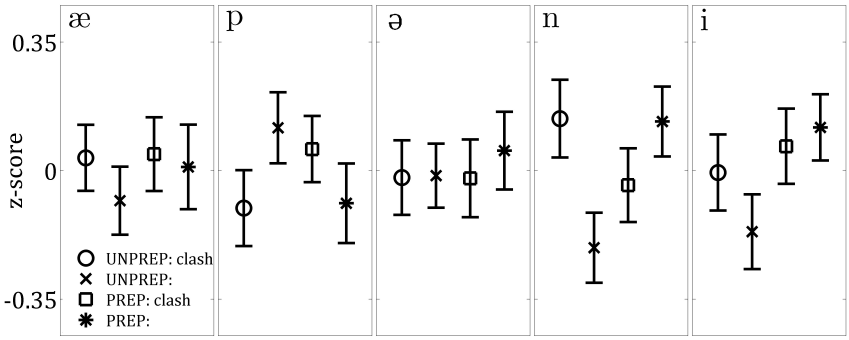

Figure 3: Mean normalized segment durations for combinations of preparation and clash.

in the secondary stressed syllable. There was indeed a marginal CLASH x PREP interaction effect on the duration of [p] (χ^2=3.30, p=0.07), which resulted in a greater increase in duration in prepared speech than unprepared speech. The segment [p] is arguably affiliated with the secondary stressed syllable by ambisyllabicity, and hence this supports the hypothesis, but the effect was not present in [ae] alone, nor in the combination of the two segments, [aep] (χ^2=1.20, p=0.27). However, as was the case for segments in the primary stressed syllable, the direction of the effect was reversed in unprepared speech, resulting in shorter [p] in the clashing context (see Figure 1). Why this would occur is somewhat mysterious. On that point, there was no CLASH x PREP interaction effect on the duration of the entire word (χ^2=1.03, p=0.31). Given the presence of such effects on the secondary stressed syllable, this implies that there is some mechanism responsible for complementarity in the durations of the primary and secondary stressed syllables.

No CLASH x PREP interaction effects were observed on any intensity measures, including relative measures that examined the ratios between the values from the primary and secondary stressed syllables. This contradicts findings from one of the previous phonetic studies discussed in section 1.4, which found clash to decrease intensity in the primary stressed syllable. Indeed, the overall pattern reveals an interesting dissociation between duration and intensity: CLASH interacted with PREP to have an effect on duration, but not intensity.

4 Discussion

The main finding, consistent with hypothesis 1, is that a prominence-reducing effect of clash was present in prepared speech only. This supports models of speech planning and production that allow for utterances to be produced before prosodic alternations have occurred, i.e., before structure is fully built. The results did not strongly indicate that a transfer of prominence occurs in prepared speech, as predicted by hypothesis 2. In what follows the results are discussed in detail along with various issues arising in their interpretation. Clash was associated with decreased duration of the primary stressed syllable in prepared speech, and unexpectedly, with increased duration of this syllable in unprepared speech. The pattern in prepared speech was predicted by hypothesis 1. It supports the idea that prosodic alternations require additional planning time to manifest; when insufficient planning time is available, such alternations are less likely to occur. Previous research has shown that prosodic structure - specifically, prosodic words - may not be fully built when an utterance is initiated (Wheeldon and Lahiri, 1997); the current findings extend this phenomenon to prosodic alternations, of which prominence-reduction in the rhythm rule context is an example. This alternation is often the parade example of prosodic alternations, but other sorts of prosodic alternations and prosodically-conditioned segmental alternations may also demonstrate similar behavior Jun (for examples, see 2005); Ito and Mester (for examples, see 2007).

The main finding of a preparation-clash interaction effect on segmental durations in the primary stressed syllable of the target was most strongly observed in [n], the onset of the syllable. The decrease was also significant when [n] and [i] were taken as a unit, but not in [i] alone. The coda [z] is expected to pattern similarly, but this cannot be resolved with the current experimental design. The duration of [z] interacts strongly with the onset segment of the following word, which in turn interacts with the stress of the word-initial syllable. Hence it cannot be known if durational patterns in the coda [z] result from stress-modulated coarticulatory influences or more global prosodic effects.

Both incremental and dynamical planning models (cf. section 1.1) predict that preparation can have an effect on the realization of prosodic structure, but the predictions arguably differ regarding the nature of the

variation. Incremental models of the sort developed by Levelt (1989) and Levelt et al. (1999) assume a prosody generator that takes as input categorical phonological representations and outputs phonetic parameters for variables such as F0, intensity, and duration. Without further mechanisms, this sort of model predicts a relatively bimodal distribution in phonetic observables associated with a given alternation: the representations that serve as input to the parameter generator are either sufficiently or insufficiently well-specified for the generator to compute an alternation, and hence one would expect some degree of bimodality in the output parameters. If planning time could be varied with high temporal resolution, then a linear increase in planning time should result in a relatively non-linear effect on parameters. In contrast, the dynamical models exhibit gradience in the representations themselves; this could allow for a somewhat more linear relation between planning-time and phonetic realization (albeit with external limits on variation for extremely short or long periods planning). It is likely, however, that stochastic influences on the time-course of planning render this prediction difficult to test experimentally.

Interestingly, the unprepared speech pattern is not exactly what was anticipated: it was predicted that clash-induced reduction of prominence would be more extensive in prepared than unprepared speech, or would occur only in prepared speech. The unexpected clash-induced increase of primary stressed syllable duration in unprepared speech does not necessarily speak against the hypothesis that prosodic alternations are more likely to apply in prepared speech than unprepared speech; instead it suggests that there is an additional planning mechanism involved, which dominates in relatively unprepared utterances. One possible explanation is that time-pressure in the unprepared condition put greater demands on perceptual or working memory processes, which could include recognizing visual stimuli, mapping from recognized pictured to lexical items, maintaining working memory of items, or accessing motor plans associated with those items. These heightened demands could have resulted in delays in how quickly motor plans associated with an upcoming word can be used for production. In support of this is the observation that, although errors in production were not very common, they occurred more frequently in unprepared than prepared speech, and more frequently between the target word and phrase-

final word than elsewhere. Extending the duration of the final syllable of the target could be a strategy for buying additional time to access working memory, or an automatic consequence of heightened demands on processing or working memory. For this explanation to make sense, the presence of initial stress on the final word (clash) would have to increase the time required for planning or accessing working memory; it is not clear why this would be the case - if anything, initial stress would be expected to facilitate lexical access (Schiller, 2006), although it is not known if there is any experimental evidence of stress-related facilitation specific to reaction-time in a production task. Another possible account (a quite speculative one) of the unprepared speech durational patterns can be derived from dynamical models of production. Stress, like phrasal boundaries, may be implemented by a production-slowing mechanism (Byrd and Saltzman, 2003) or alternatively, by mutually inhibitory stress planning systems which heighten activation of associated gestural planning systems, thereby extending the duration of those gestures (Tilsen, 2011a,b). In the latter case, nearby stresses in unprepared speech may overlap and imbue gestures with additional duration. In prepared speech the stress planning systems would have more time to mutually inhibit one another and hence this "stress overlap effect" would be diminished.

The manipulation of preparation was a key aspect of the experimental design. It should be clarified that utterances produced in the "prepared" and "unprepared" conditions are both prepared to some extent - the difference is a matter of relative preparation. In order to control preparation, subjects were instructed to subvocally rehearse the utterance one time before producing it. One possible objection is that referring to the controlled variable as "preparation" misses the true source of the effect: the subvocal rehearsal performed by speakers in the prepared condition. It is conceivable that the source of the effect is not due to preparation time, but rather to the performance of a subvocal rehearsal prior to utterance initiation. However, it is quite sensible to equate the rehearsal itself with "preparation" or planning. Indeed, it is not clear exactly what preparation would be if not some form of unvocalized rehearsal of an utterance. Whether the concepts of rehearsal and preparation can be operationally distinguished in an experimental paradigm remains an open question.

The results provided quite weak and ambiguous support for hypothesis 2, although in the absence of intonational analysis they cannot be interpreted as conclusive. Derived from transfer-based phonological analyses of the rhythm rule, hypothesis 2 predicted that in a clashing context, the secondary stressed syllable would exhibit an increase in prominence due to a transfer of prominence from the primary stress. The duration of [p] was marginally increased by clash in prepared speech, but [ae] duration was unaffected. The segment [p] could be viewed as associated with the secondary stressed syllable, via ambisyllabicity, although it is more typically viewed as the onset of the unstressed medial syllable in "Japanese". As with the primary stress onset [n], the pattern with [p] was reversed in unprepared speech: clash shortened [p]. In other words, [p] and [ni] exhibit opposing patterns of durational variation across conditions, revealing a somewhat striking complementarity between segmental durations in the final syllable and preceding syllable. One possible explanation for this is that speakers exert global control on individual word durations, and hence durational effects on the primary stress are compensated for with opposing effects earlier in the word.

It is likely that analysis of pitch and phrasal boundary accents may shed further light on the source of experimental effects. Treatments of pitch accent clash from Bolinger (1965) and Shattuck Hufnagel et al. (1994) predict that clash should decrease F0 in the primary stress of the phrase-medial target word. An alternative possibility is that clash is repaired by augmentation of a prosodic boundary between the target word and the phrase-final world, which is partly consistent with the silent beat analysis of Selkirk (1984). An augmented boundary would result in the production of an augmented pitch accent on the primary stressed syllable of the target word. However, this would also predict that clash should have lengthened duration of this syllable, at least in the prepared responses, which was not the case. Hence the idea that clash augmented the word-final boundary is not very convincing.

Finally, a dissociation between effects on duration and intensity was observed: while duration was differently affected by the interaction of clash and preparation, only a marginal main effect of clash was observed on intensity, which resulted in an increase of intensity in the vowel in the secondary stressed syllable. It should be noted that the task design is not particularly well-suited for investigating more global prosodic parame-

ters such as F0 and intensity. Subjects produced many repetitions of similar phrases during the experiment, and they did so in a communicative context that lacked a listener. Primary functions of F0 and intensity in English include the signaling of turn-taking, informational structure, and pragmatic context. Both the repetitive nature of the task and the absence of a listener may have led subjects to under-utilize these variables. It may be the case that F0 and intensity are not manipulated extensively enough to reveal differential effects of clash between preparation conditions.

5 Conclusion

The main finding of this experiment is that the effect of clash on segmental duration differs depending upon the extent to which an utterance has been prepared. Previous studies have almost exclusively examined prepared utterances, and have found some evidence for clash-induced reduction of the primary stress in the rhythm rule context. The present study suggests that this prominence reduction does not occur in relatively unprepared speech. The theoretical implication of this result is that prosodic structure can be incompletely built upon the initiation of an utterance, and prosodic alternations are less likely to occur in speech that is less extensively prepared. This is consistent with current models of speech planning and production, but is not obvious from static phonological representations of prosodic structure.

Another finding is that durational effects of clash were not observed on the preceding secondary stressed syllable, in either prepared or unprepared responses. This is consistent with previous phonetic studies which have generally failed to observe evidence for a "transfer" of prominence in the rhythm rule context, and it supports prominence-reduction based analyses of the prosodic alternation. However, it is possible that analysis of pitch accentual patterns may reveal a transfer of pitch accent in this context, albeit without corresponding durational changes.

An important contribution of this study is the development of an experimental paradigm suitable for investigating the effects of planning time on prosodic structure building and prosodic alternations. Preparation

was manipulated by task instruction and stimuli were presented visually to elicit formulaic three-word noun phrases. These methodological devices provide a relatively simple framework for conducting studies of prosodic planning. Moreover, future studies which aim to understand the effects of prosodic structure on speech should take into account the extent to which speakers prepare their utterances, since preparation can have potentially significant effects on phonetic patterns. Experimental manipulation of preparation can provide a potentially rich source of data to inform models of speech planning and production.

6 Acknowledgements

This research was supported by NIH/NIDCD grants #R01-DC006435, #R01-DC003172-14, and #R01-DC008780-04. I would like to thank Louis Goldstein, Dani Byrd, and Rachel Walker for advice in the design and analysis of this experiment. Thanks to Ed Holsinger and Ben Parrell for lively discussions of linear mixed effects regression, which aided the analyses of these data.

References

Al-Mozainy, H. Q., Bley-Vroman, R., and McCarthy, J. J. (1985). Stress shift and metrical structure. *Linguistic Inquiry*, 16(1):135–144.

Beckman, M. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P. A., editor, *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, pages 1–33. Cambridge University Press, Cambridge U.K.

Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S. A., editor, *Prosodic Typology: The Phonology of Intonation and Phrasing*, pages 9–54. Oxford University Press, Oxford.

Bing, J. M. (1980). Linguistic rhythm and grammatical structure in Afghan Persian. *Linguistic Inquiry*, 11(3):437–463.

Bolinger, D. (1958). A theory of pitch accent in English. *Word*, 14(2-3):119–149.

Bolinger, D. (1965). Pitch accent and sentence rhythm. In Abe, I. and Kanekiyo, T., editors, *Accent, Morpheme, Order*, pages 139–180. Hokuo, Tokyo.

Bolinger, D. (1985). Two views of accent. *Journal of Linguistics*, 21(1):79–123.

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

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57(1):3–16.

Byrd, D. and Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26:173–199.

Byrd, D. and Saltzman, E. L. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2):149 – 180.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* Harper & Row, New York.

Churchyard, H. (1999). *Topics in Tiberian Biblical Hebrew metrical phonology and prosodics.* PhD thesis, Department of Linguistics, University of Texas, Austin.

Cooper, W. E. and Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, 25(3):369–384.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1:126–152.

Grabe, E. and Warren, P. (1995). Stress Shift: Do speakers do it or listeners hear it? In Connell, B., editor, *Laboratory Phonology IV. Phonology and Phonetic Evidence*, pages 95–110. Oxford University Press, Oxford.

Gussenhoven, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology*, 8(1):1–35.

Gussenhoven, C. (2011). Sentential prominence in English. In van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*, pages 2780–2806. Wiley-Blackwell, Malden M.A. & Oxford.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies.* University of Chicago Press, Chicago.

Hayes, B. and Puppel, S. (1985). On the rhythm rule in Polish. In van der Hulst, H. and Smith, N., editors, *Advances in Nonlinear Phonology*, pages 59–81. Foris Publications, Dordrecht.

Horne, M. (1990). Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, 28 (5):959–982.

Ito, J. and Mester, A. (2007). Prosodic adjunction in Japanese compounds. In *Proceedings of Formal Approaches to Japanese Linguistics 4*, pages 97–112.

Jun, S. A. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, Oxford.

Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2):347–365.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10(3):421–441.

Klapp, S. T. (1995). Motor response programming during simple choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5):1015.

Klapp, S. T. (2003). Reaction time analysis of two types of motor preparation for speech articulation: Action as a sequence of chunks. *Journal of Motor Behavior*, 35(2):135–150.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3):129–140.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* MIT Press, Cambridge M.A.

Levelt, W. J. M., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22:1–75.

Liberman, M. (1975). *The Intonation System of English.* PhD thesis, MIT, Boston.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.

Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, 38(1-3):84–105.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology.* Foris Publications, Dordrecht.

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

Saltzman, E. L., Nam, H., Krivokapic, J., and Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Barbosa, P. A., Madureira, S., and Reis, C., editors, *Proceedings of the 4th Conference on Speech Prosody, Campinas*, pages 175–184.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, 1112(1):201–212.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press, Cambridge M.A.

Shattuck Hufnagel, S., Ostendorf, M., and Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*, 22:357–388.

Sternberg, S., Knoll, R. L., Monsell, S., and Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45 (2-4):175–197.

Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In Stelmach, G. E., editor, *Information Processing in Motor Control and Learning*, pages 117–152. Academic Press, New York.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, 33:839–879.

Tilsen, S. (2011a). Effects of syllable stress on articulatory planning observed in a stop-signal experiment. *Journal of Phonetics*, 39(4):642–659.

Tilsen, S. (2011b). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, 2:642–659.

Tilsen, S. and Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124(2):34–39.

Vanderslice, R. and Ladefoged, P. (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language*, 48(4):819–838.

Vogel, I., Bunnell, H., and Hoskins, S. (1995). The phonology and phonetics of the Rhythm Rule. In Connell, B., editor, *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence*, pages 111–127. Cambridge University Press, Cambridge U.K.

Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.