RUNNING HEAD: Fluency effects in human language

Fluency effects in human language: the integration of automatic and intentional mechanisms of acoustic variation in speech

Jennifer E. Arnold

Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

Address all correspondence to: Jennifer E. Arnold, UNC Chapel Hill, Dept. of Psychology and Neuroscience, Davie Hall #337B, CB #3270, Chapel Hill, NC 27599-3270, Email: jarnold@email.unc.edu

ABSTRACT

 This paper examines one aspect of human language: variation in acoustic prominence. This variation is typically explained in terms of its contribution to the communicative goal of language, in that it helps signal information status. I review evidence that acoustic variation is also influenced by the cognitive processes involved in language production, such that more difficult production situations result in slowed pronunciation, which contributes to the impression of acoustic prominence. Critically, production difficulty correlates with the information status of the word being uttered, in that reference to new information is harder than reference to given information. This highlights one way in which the social/communicative goals of language are subserved by correlated processes that are relatively automatic, but unintentional and unrelated to the communicative value of the signal.

Seyfarth and Cheney (this volume) argue that the hallmark of human language is the use of a discrete, rule-governed system to achieve social communicative goals. People generally believe that when they talk, they do so intentionally. That is, people say things because they mean them, and other people use those words and sentences to infer the speaker’s meaning. Seyfarth and Cheney go on to argue that nonhuman communicative systems share some of these properties, supporting the view that both communicative systems evolved out of similar social systems. In a nutshell, they argue that baboon communication is an intentional and socially functional system. This challenges the perception that animal communication differs categorically from human language, and points toward evolutionary continuity. This argument also bridges the contrast between highly flexible and productive human language systems and relatively inflexible and innate animal communication systems (Cheney & Seyfarth, 1997).

In this essay I argue that the comparison between human and nonhuman communictive systems also requires a closer look at human language. In particular, we need to consider the ways in which the intentional aspects of language are rooted in less intentional, automatic aspects of cognition and behavior. I will argue that unintentional behavior plays a systematic role in the types of human signals that are used communicatively and intentionally. Here I examine this pattern with respect to the ways in which speech variation reflects the information status of the utterance.

When people speak, they exhibit striking variability in the pronunciation of their words. For example, the same speaker might produce the sentence “this food is great” by emphasizing “food” on one occasion “This *food* is great”, but emphasizing the determiner on another *This* food is great”. This illustrates one of the most ubiquitous domains of linguistic variation, which is the degree of intelligibility, or acoustic prominence of a word. This kind of variation contributes to a dimension of language known as **prosody**, which referes to the timing, pitch, rhythmic, and acoustic properties of speech. Words vary in acoustic properties such as duration (longer vs. shorter), pitch (e.g. high vs. low, or rising vs. falling), and intensity (loud vs. quiet). Speakers also vary in how clearly they pronounce their words, sometimes attenuating or dropping certain phonemes (*Imuna* vs. *I’m going to*), and varying the degree to which their vowels are distinguishable (Bradlow, Torretta, & Pisoni, 1996). In all these dimensions, pronunciations can vary from more **reduced** expressions (short, quiet, unintelligible, low pitch or little pitch movement) to more **prominent** expressions (long, loud, intellible, higher pitch or more pitch movement).

This variation is interesting because it relates in systematic ways to the message that the speaker is trying to communicate. In particular, it tends to reflect the speaker’s assumptions about what their addressee knows or is attending to, or the recent conversational context. This is known as information status.

**Information status effects on acoustic prominence vs. acoustic reduction**

Imagine that Elise says to Jason “The new lab computer isn’t working.” At this point, she can assume that by mentioning the new computer, she and Jason can both assume that they share knowledge of the computer, and additionally are currently attending to it. Previously evoked information like this is termed “given” (or “old”) information, and contrasts with new information. Given that Elise introduced the computer into the conversation, they can probably also assume that it will be important in the upcoming conversation, making further mention of the computer relatively predictable. Both givenness and predictability are components of the computer’s information status in the discourse (e.g., Chafe, 1976; Prince, 1981).

 There is substantial evidence that information status guides both the way we formulate our ideas in words, and the way we understand what other people mean. For example, it sounds more natural to say “That’s my dog. My dog chased the cat”, than “That’s my dog. The cat was chased by my dog”. This demonstrates that speakers are more likely to start their sentence with given information than with new information (Arnold et al., 2000; Bock & Irwin, 1981). Listeners also assume that words that come later in the utterance refer to something that is discourse-new (Arnold & Lao, 2008).

 Information status also guides acoustic prominence vs. acoustic reduction. For example, if Elise says “Where is Sandy?” Jason is likely to then pronounce “Sandy” with an unaccented, acoustically reduced expression, which reflects her status as given and attended. This reflects one of the most reliable effects in acoustic prominence: repeated mention. If the same person or object is mentioned twice in a conversation, the second mention tends to be shorter in duration, lower in pitch, involve less pitch movement, and quieter (Ladd, 1996). However, simple re-mention is not enough. For example, Terken and Hirschberg (1994) asked speakers to describe object movements like “The ball touches the cone, the cone touches the ball.” The words “cone” and “ball” in the second sentence were not reduced, unless they appeared in the same roles as they did in the first sentence (as for the second “ball” in, “The ball touches the cone, the ball touches the square”). This effect of parallelism may be related to the fact that parallel mention is more expected than nonparallel mention of given information (Arnold, 1998), suggesting that reduction is related to the continuity of the discourse. Relatedly, Fowler (1988) has claimed that reduction does not occur unless the word refers to the same referent.

Speakers also tend to use acoustic reduction for information that is predictable from the context. In a classic experiment, Lieberman (1963) demonstrated that speakers pronounced the word “nine” with greater emphasis in an unconstraining context (“The next number you will hear is nine”) compared to a context that made the word predictable (“A stitch in time saves nine”). More recent work has shown that in running speech, words tended be shorter and more unintelligible if they are statistically likely to co-occur with other words in their context (Bell et al., 2009; Jurafsky et al., 2001), or if they are in a syntactically probable construction (Gahl & Garnsey, 2004).

 While reduced forms are used for given and predictable information, speakers use fuller, accented forms for information that is new or less predictable. Likewise, prominent (e.g., emphatic-sounding) pronunciations are used when the speaker wishes to communicate contrast. For example, Jason might say “SANDY is in the lab, but I don’t know where KATHRYN is,” (Ito & Speer, 2008).

**Is acoustic prominence produced intentionally?**

 As the preceding section shows, information status effects on acoustic prominence are ubiquitous and well established. Yet there is less consensus about why these effects occur. Understanding why is relevant to questions about the continuity of language evolution, and more broadly the functional nature of communication systems. Seyfarth and Cheney (citing Clark, 1996) point out that human language serves specific functions, for example in that it is typically used for specific social purposes, and always involves both speaker’s meaning and addressee’s understanding. This view fits well with the traditional view on information-status effects. For example, Grice (1975) proposed that successful human communication depends on speakers and listeners following certain maxims, and assuming that they can interpret other’s communications in light of these maxims. One such maxim considers quantity, and suggests that speakers should say just as much as is needed for communication, but not too much. In keeping with this, speakers tend to provide an explicit bottom-up signal in the form of greater acoustic prominence (which confers greater intelligibility) when information is not recoverable from the context – i.e., when it is new or unpredictable.

This view represents a functional (or grammatical) explanation of information-status effects. This type of mechanism serves to meet the social/communicative function of language, and it assumes that the acoustic form is chosen because it corresponds to the speaker’s intended meaning. In this sense, it is a part of the intentional purposes of speaking, even though speakers do not need to be aware of the mechanisms by which linguistic forms are selected.

The functional explanation predicts that listeners should be able to use acoustic variation to help them identify the right referent, and indeed evidence suggests that they can (Arnold, 2008; Dahan, Chambers, and Tanenhaus 2002). For example, Arnold (2008) tracked eyegaze of both adults and children as they followed instructions to move pictures on a computer screen. Subjects viewed a set of four pictures, including two with similar-sounding names (e.g., bagel, bacon). They heard an instruction mentioning one of them “Put the bacon on the square”. The second sentence was the critical one, mentioning the target with either an accented or unaccented expression “Now put the BACON on the triangle.” Critically, at the onset of the target expression “Ba-…”, the input was ambiguous, as it could potentially be the start of either *bacon* or *bagel.* When the word was unaccented, listeners had a tendency to look at the unmentioned (discourse-given) object, but when it was accented, this bias disappeared. Thus, prosody helped direct the listener to the more contextually likely referent.

The functional explanation also accords with the fact that people have strong intuitions about the “right” way to say something. For example, it just sounds “wrong” to repeatedly emphasize a referent in a story:“ALEX went to the museum, and ALEX bought a ticket, and ALEX saw the show”*.*

 However, the fact that acoustic prominence plays a useful role in communication does not mean that it was produced intentionally. Seyfarth and Cheney (1992) discuss the question of how we (as scientists) can know whether an animal’s call is produced intentionally or not. They tell the story of the tennis player Jimmy Connors, who was in the habit of grunting loudly as he hit the tennis ball. When officials complained, he claimed that the grunt was unintentional, just a side-effect of physical effort. Without any way to assess this claim, they could not fault him for it.

Similarly, it is possible that the cognitive mechanism that leads speakers to produce variation in acoustic prominence is not directly influenced by the social goal of communication. An alternate possibility is that acoustic variation results from relatively automatic mechanisms, but in such a way that it correlates systematically with information status. If so, perhaps listeners simply learn to make use of acoustic variation as a correlational cue.

Indeed, we know that humans (and other animals) are extraordinarily good at picking up on patterns in their environment, and can use them to draw inferences and make predictions. One example of this comes from evidence that listeners can even use speech disfluency to help them anticipate the speaker’s meaning. It turns out that speakers are more likely to be disfluent when they are referring to something new than something given, consistent with the idea that new references require more planning and are cognitively more difficult (Arnold & Tanenhaus, 2011). It is implausible that disfluency is produced intentionally as a signal about an upcoming reference to something new.[[1]](#footnote-1) Yet comprehenders can still make use of the systematic relationship between disfluency and discourse newness. In one study (Arnold, Tanenhaus, Altmann, & Fagnano, 2004), we demonstrated that disfluency led to a bias toward discourse-new objects. For example, following “Put the grapes below the candle” (which made the candle “given”), listeners expected the next sentence to include a reference to the candle if the speech was fluent. However, if speech was disfluent, they expected reference to an unmentioned object. Other work has suggested that disfluency leads to a general bias toward things that are difficult to name (Arnold, Hudson Kam, & Tanenhaus, 2007; Heller, Arnold, Klein, & Tanenhaus, 2014).

Thus, we must consider the possibility that acoustic prominence may result from other, more automatic, constraints on the language production process, and its communicative function derives from the listener’s ability to make use of any cues that are available. Indeed, there is extensive evidence that acoustic variation is related to the ease and fluency with which speakers can produce their utterances.

**Acoustic prominence is related to speech production fluency**

 Human language is a complex system, and scholars agree that turning thoughts into words requires manipulating information at numerous levels. For example, speakers need to generate the concepetual structure, think of the words and syntactic structures they need, generate the phonologial form for the utterance, and program the motor movements needed to articulate it (e.g., Levelt 1989). Critically, each of these processes takes time and cognitive resources. Thus, anything that changes the ease of planning an utterance might change the speed and fluency with which speakers can produce it. The critical question for our current purposes is whether this changes the prosodic form of a word.

 Evidence suggests that planning does indeed affect reference form. Support for this idea comes primarily from evidence that the duration of a word is related to various measures of word planning difficulty (Arnold & Watson, 2015; Gillespie, 2011; Kahn & Arnold, 2012; Zerkle & Arnold, under review). This makes sense, in that planning difficulty likely changes the timing of both planning and articulation – thus, timing variation affects the timing component of prosody, duration.

One effect of planning emerges from the observation that speech planning needs to precede actual articulation. However, speakers have a choice: they can either completely pre-plan an utterance before saying it, or they can plan as they go along. The “on the fly” option still requires pre-planning, but it means that the speaker plans word x while they are uttering words x-1 and earlier. As a result, difficult words tend to result in slowing on the **previous** word.

For example, Christodoulou (2012) asked participants to name pairs of pictures, for example “Skunk hand”, or “Skunk deer”. The critical manipulation was the frequency of the second word, which was either high (“hand”) or relatively low (“deer”). Frequency is well known to influence the speed with which speakers retrieve words, and thus represents a manipulation of relative planning difficulty. Subjects consistently produced a slower “Skunk” preceding a low-frequency than preceding a high-frequency word. In a series of experiments using eyetracking, Christodoulou critically provided support for the role of utterance planning in this finding. He found that the timing with which participants fixated the second picture (hand or deer) mattered, such that earlier fixations led to shorter first-word durations. The first fixation on the second picture was taken to be a measure of when they began planning the second word (including conceptual processing), suggesting that the duration of word 1 is sensitive to the speaker’s degree of readiness to produce word 2. In experiments 3 and 4, results critically demonstrated that the timing of planning word 2 modulated the effect of its duration on word 1. If speakers looked at object 2 **before** they began uttering word 1, there was a strong effect of word 2 frequency, such that word 1 was shorter if word 2 was frequent. However, if speakers began uttering word 1 before fixating on 2, they used a uniformly long duration, regardless of word 2 frequency.

Christodolou’s work illustrates how speech planning can influence the timing of the speech regions that precede the planned unit. Other work shows that ease of planning a speech unit affects the pronunciation of that unit itself. One well-established effect is that words tend to be shorter if they are more probable. A simple measure of probability is frequency, which is a kind of context-free probability – i.e., the likelihood of a word in the language overall. Zipf (1929) established a classic effect by which frequent words (e.g. “car”) tend to be shorter than infrequent words (e.g., “automobile”; see also Piantadosi, Tily, & Gibson, 2011). This finding extends to the pronunciation of a particular token. For example, homophones like “time” and “thyme” tend to differ in their duration, such that the less frequent member tends to be pronounced with longer durations (Gahl, 2008).

 Similarly, word duration is sensitive to the probability of a word, contingent on both the preceding and following words. That is, the duration of word n depends on how predictable it is based on knowing word n+1, and also on how predictable it is based on knowing word n-1. A good example of this comes from names (Christodoulou, 2012): the word “Ford” is relatively probable following “Harrison” (forward predictability), while the word “Burt” is relatively probable when it precedes “Reynolds” (backward predictability). Bell et al. (2009) found that both preceding and following context matter, such that higher probability words tend to be pronounced more quickly, and with less clarity.

 While researchers disagree about the origin of these probability effects, a plausible explanation is that word probability facilitates production: easy-to-produce words are retrieved more quickly, and allowing the speaker to plan the following word and produce the two as a unit (Arnold & Watson, 2015). This results in fluent, connected speech.

 Reduction effects extend beyond just the predictability of words, and also include the predictability of actions. Watson, Arnold, and Tanenhaus (2008) demonstrated this by having pairs of participants play a verbal game of Tic Tac Toe. They informed each other of their moves, saying things like “My marker goes on nine,” where the number deonted the grid location. If the move was a winning move, the location was completely predictable to both players, due to the fact that the goal of the game encourages players to put their marker in a space next to two other markers, if possible. Likewise, a block move was highly predictable. We found that speakers pronounced the number word (e.g., “nine”) more quickly when the move was predictable (i.e., when it was a win/block) than when it was not. In this task, the goal of winning made a move predictable, possibly enabling subjects to begin planning their utterance earlier, pre-preparing the sounds in order to meet fluency. It may also have decreased uncertainty about the move, such that they did not need to slow down to monitor the correctness of their move.

 Further evidence that speech difficulty affects prosody comes from analyses of the conditions that lead to disfluency. Speakers often fail to achieve the goal of fluent speech delivery: they hesitate, repeat their words, restart their sentence, or produce filler words like “uh” or “um”. Bell et al. (2003) did a large-scale analysis of the pronunciation of function words like “and”, “the”, and “of” in a speech corpus, and found that words immediately preceding and following disfluent elements tended to have fuller pronunciations. This emerged both in the duration of the pronunciation, and in the speaker’s choice between the full vowel form (e.g., “thiy” (rhyming with tree) vs. “thuh” for “the”). This suggests that speech difficulty is associated with non-reduced pronunciations.

 One proposal for why fluency affects prosody focuses on the sound representations for each word. Watson, Buxo-Lugo, and Simmons (2015) propose that there is a relationship between the ease of retrieving the phonological code for a word and the time it takes to pronounce it. Most models propose that word production involves separate stages for the selection of the word, and the selection of the phonological code (e.g., Dell, 1986). Watson et al. argue that speakers may begin to articulate a word before they have finished selecting the phonological code for the end of the word. If so, lengthening the pronunciation would provide the time needed to encode the sound structure. They present data from both human participants and a computational model to support this proposal.

Waton et al.’s proposal is also consistent with evidence that predictable words are shortened. When a word is predictable, the speaker can begin planning it earlier. This would result in the earlier selection of all aspects of the word, including its phonological code.

A related proposal is that acoustic reduction occurs when the word is pre-activated (Kahn & Arnold, 2012). Under this view, reduction is more likely when the speaker is already thinking about both the concept and the linguistic word itself, compared with just the concept. Kahn and Arnold demonstrated exactly this pattern, using a task where speakers described moving objects (e.g., “The accordian rotates”). When the accordian was predictable (based on a pre-speech cue), speakers produced the word more quickly. When speakers also heard the actual word “accordian” before their response, the duration was even shorter.

 Another possibility is that highly frequent words are easier to produce because the articulatory processes are routinized (Bybee & Hopper, 2001). This means that highly practiced sequences are produced more quickly, so it accounts for effects that are learned over time, like frequency or co-occurrence probabilities. However, a variation on this would be needed to account for facilitation effects that arise from the current situation, such as repeated-mention effects.

 Another possibility is that planning difficulty of one word can impact the timecourse of planning subsequent words. Consider the sentence “The Venetian vase goes next to the window.” If the speaker has momentary difficulty retrieving the word “Venetian”, the planning system will require the support of all cognitive resources to successfully retrieve it. This leaves few resources available to concurrently plan the following word, “vase”. This requires the speaker to slow down on “Venetian” while they plan the next word. This account predicts that facilitation effects are most likely to occur in the context of multi-word utterances, and in a task that encourages on-the-fly planning, as opposed to pre-planned speech.

 Critically, the fluency effects reported here are somewhat independent of the social/communicative function of language. In one sense they are not completely dissociated, because speech difficulty only occurs when the speaker is attempting to generate an utterance, and this activity is directed by the need to satisfy a social goal. However, fluency effects themselves are not driven by communicative goals. The speaker might intend to introduce a new concept “Look at the Venetian vase,” but this creates difficulty, which is the direct cause of duration variation, and perhaps disfluency (“Look at theee uh Venetian vase.”)

Consistent with this, several studies have found evidence that the intelligibility and duration of spoken words is unaffected by what the listener knows. These findings contradict a hypothesis that acoustic variation is driven by audience design – that is, speakers produce reduced forms precisely when comprehension is facilitated, for example because the word is repeated or predictable (Lindblom, 1990). Instead, many studies have found that variation in the listener’s knowledge – for example, whether the listener heard the first mention of the word – has little to no impact on the speaker’s tendency to reduce a second mention of the word (Bard & Aylett, 2004; Bard et al., 2000; Kahn & Arnold, 2015). However, feedback from the listener can affect fluency of utterance planning and production (Arnold, Kahn, & Pancani, 2012).

 In sum, extensive evidence suggests that the variation in speech pronunciation is systematically related to the process of production. When a word or phrase is difficult to plan, speakers slow down, and become less fluent. They also use phonologically more explicit forms, like full vowels, and resist simplifying consonant clusters. This suggests that the psychological mechanisms of speech production have a direct impact on linguistic form.

**Why fluency effects matter: Fluency and information status**

 The premise behind this essay is that it is worthwhile considering the degree to which human language involves relatively automatic mechanisms that are not directly managed by the speaker’s social/communicative goals. In the previous section I demonstrated that the ease of planning an utterance has consequences for linguistic form. These planning effects are most naturally viewed as a side-effect of constructing an utterance, suggesting that processing difficulty itself modulates linguistic form.

Yet such effects are not surprising by themselves. What is more interesting is that processing difficulty and speech fluency are systematically related to the kinds of information-status categories that have been proposed to underlie linguistic form for functional reasons.

For example, consider the contrast between “given” (previously mentioned or otherwise evoked) and “new” information. New information is likely to be more difficult to talk about than given information, because it requires more cognitive resources to build a mental representation of the concept, and the words should be harder to retrieve. Speakers may also allocate more cognitive resources to monitoring words that represent new information, perhaps to check the correctness of the word or message. Speakers may also need more time to construct the phonological representation for new words. This predicts that speakers should slow down for new words, both before and during the word. Indeed, this pattern of reduction for repeated words has been widely reported (Bell et al., 2009; Brown, 1983; Fowler & Housum, 1987; Halliday, 1967).

Yet the processing-based explanation contrasts with the traditional explanation for acoustic variation in spoken language. It is widely accepted that some languages (like English) mark information status via representations in the domain of prosody (e.g., Pierrehumbert & Hirschberg, 1990). One aspect of prosody is the placement of a pitch accent, which corresponds to greater acoustic prominence of the word. This grammatical approach can be classified as a part of the functional approach described above, in that it suggests that particular prosodic forms are licensed by grammatical or pragmatic rules, and in particular that prosody is used to mark information status. That is, prosody is a tool for achieving a communicative goal.

For example, Schwarzchild (1999) argues that the grammar includes the constraint that speakers should avoid accenting given information. Thus, a listener can infer that an unaccented word is given, while an accented word is probably new (for other grammatical approaches, see Gussenhoven, 1983; Selkirk, 1996). While these theories do not specify the mechanisms by which grammar shapes linguistic form, it seems plausible that they should do so directly, such that a particular intended message leads to the selection of the appropriate linguistic form (Arnold, in press; Arnold & Watson, 2015; Kahn & Arnold, 2012).

Thus, it is notable that information status can have two different types of influences on reference form. The functional (aka “intentional”) influence is through the grammar: reduced forms are appropriate in some contexts, such as when the referent is given. The processing (aka “automatic”) influence results from the relative ease of production processing that occurs from given or salient information, which leads to shorter and more attenuated pronunciations. Critically, both types of mechanisms have similar effects.

While the correlation between functional and processing accounts is a challenge for researchers, there is reason to believe that processing effects are intertwined with the functional use of prosody. First, timing variation is part and parcel of the accenting categories that are relevant to grammatical theories of prosody. When a word is accented, it affects several acoustic properties, including pitch, pitch movement, intensity, and the relative duration of the word with respect to other words in the utterance (Ladd, 1996). For example, Breen et al. (2010) examined the pronunciation of sentences like “Damon fried an omelet”, and manipulated the location of focus between Damon, fried, and omelet. They found that focus led speakers to produce words with longer durations, larger f0 excursions, greater intensity, and longer subsequent pauses, compared to when it was not focused. Thus, duration contributes to the expression of information status. Since duration is the dimension of prosody that is most likely to be affected by processing difficulty, this opens the door for speech difficulty itself to affect acoustic prominence.

Moreover, there is evidence that duration variation can influence speech comprehension as well. Using the same experimental paradigm as Arnold (2008), Arnold, Pancani, & Rosa (2015) tracked listeners’ eye movements as they followed instructions like “Put the bagel on the square. Now put the bacon on the circle”, and compared acoustically prominent word “Now put the BACON…” with reduced ones “Now put the bacon…” Critically, we also manipulated the overall fluency and speech rate. Half the subjects were told that the speaker was distracted while speaking, and all of the sentences were spoken in a slow and halting manner. This had the effect of slowing all the words in the sentence, including the target word. The question we asked was whether the overall slowing would change the perception of the target word. If only relative prominence matters (as the grammatical approach would suggest), then it should not. By contrast, we found that in the distracted condition, listeners had a stronger bias toward the discourse-new referent than in the undistracted condition. This occurred despite the fact that the slowing was clearly due to the speaker’s distracted, disfluent delivery, and not the information status itself. On the other hand, listeners were still able to use pitch variation to distinguish the reduced and unreduced conditions, even in the distracted condition. This suggests that acoustic prominence effects cannot be explained entirely in terms of duration variation. Rather, the listener’s perception of acoustic prominence stems from all sources of information, including slowing that stems from speaker difficulty.

**Conclusions**

The research reviewed here focuses on the fact that human speech involves variation in pronunciation. Sometimes speakers use acoustially prominent pronunciations, which are longer, louder, and display greater pitch movement. Sometimes speakers use acoustically reduced pronunciations, which are quieter, softer, and often phonologically simpler.

The classic explanation for this variation focuses on the communicative goal of language: speakers use particular forms to signal the information status of their intended message. Reduced forms generally reflect given or unfocused information, while new forms are used for new or unpredictable information. This explanation matches the view that human language (unlike animal communication systems) is characterized as a complex, rule-governed system of communication.

There is no question that languages include grammatical constraints on prosody. Yet in contrast with the traditional view, I have argued for a more complex system. Acoustic variation is not just the result of prosodic marking, but it can also result as a side-effect of the cognitive processes necessary to produce an utterance. Thus, a reduced pronunciation may not be selected entirely on the basis of the speaker’s communicative intention. Nevertheless, processing-based variation impacts the listener’s interpretation. Thus, fluency and processing effects are an integral part of prosody.

This view raises new possibilities for identifying continuities between human and animal language communication systems. Seyfarth and Cheney argue that nonhuman primate communication serves many of the same social goals as human language. On the flip side, I argue that human language incorporates some relatively automatic processes – a type of mechanism that is often attributed to animal systems. Fluency effects may not be designed specifically for a social purpose, but they correlate with more purposeful grammatical effects. Thus, fluency contributes to successful communication, the ultimate goal of language use.

AUTHOR NOTE

This work was partly supported by NSF grant 0745627 to J. Arnold.

REFERENCES

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, 108(1), 69–99. doi:10.1016/j.cognition.2008.01.001

Arnold, J. E. & Tanenhaus, M. K. (2011). Disfluency isn’t just um and uh: the role of prosody in the comprehension of disfluency. In Gibson, E., and Perlmutter, N. (Eds) *The processing and acquisition of reference*, MIT Press.

Arnold, J. E. & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: performance matters. Language, Cognition, and Neuroscience. DOI:10.1080/01690965.2013.840733

Arnold, J. E., & Lao, S. C. (2008). Put in last position something previously unmentioned: Word order effects on referential expectancy and reference comprehension. *Language And Cognitive Processes, 23*(2), 282-295. doi:10.1080/01690960701536805

Arnold, J. E. (in press, 2016 online). Explicit and Emergent Mechanisms of Information Status. *Topics In Cognitive Science*. Doi: 10.1111/tops.12220

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal Of Memory And Language, 51*(1), 55-70. doi:10.1016/j.jml.2004.03.006

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 33*(5), 914-930. doi:10.1037/0278-7393.33.5.914

Arnold, J.E., Kahn, J.M., & Pancani, G. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin and Review*, 19, 505-512. DOI 10.3758/s13423-012-0233-y

Arnold, J. E., Tanenhaus, M. K. , Altmann, R., & Fagnano, M. (2004). The Old and Thee, uh, New. *Psychological Science*, 578-582.

Arnold, J. E., Pancani, G. C., & Rosa, E. C. (2015). Acoustic prominence is perceived differently for fluent and distracted speakers. Unpublished manuscript, University of North Carolina at Chapel Hill.

Bard, E. G., and M. P. Aylett 2004. Referential form, word duration, and modeling the listener in spoken dialogue. *Approaches to studying world-situated language use: bridging the language-as-product and language-as-action traditions*, ed. by J. C. Trueswell and M. K. Tanenhaus, 173–91. Cambridge, MA: MIT Press.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1–22. doi:10.1006/jmla.1999.2667

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92–111. doi:10.1016/j.jml.2008.06.003

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C.Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America,* 113, 1001–1024. doi:10.1121/1.1534836

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 19, 467–484.

Bradlow, A. R., Torretta, G., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255–272.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25, 1044–1098. doi:10.1080/01690965. 2010.504378

Brown, G. (1983). Prosodic structure and the given/new distinction. In A.Cutler, & D. R. Ladd (Eds.), Prosody: Models and measurements (pp. 67–77). New York: Springer-Verlag.

Bybee, J., & Hopper, P. (2001). Introduction to frequency and emergence of linguistic structure. In J. Bybee & P. Hopper (Eds.), Frequency and the emergence of linguistic structure (pp. 1–24). Amsterdam: Benjamin.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and topic, ed. by C. N. Li, 25–56. New York: Academic Press Inc.

Cheney, D. L., & Seyfarth, R. M. (1997). Animals don’t have language. Tanner Lectures on Human Values, delivered at Cambridge University, March 10-12, 1997.

Christodoulou, A. (2012). Variation in Word Duration and Planning. Ph.D. thesis, University of North Carolina at Chapel Hill.

Clark H. H., & Fox Tree J. E. 2002. Using uh and um in spontaneous speaking. Cognition 84: 73–111.

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. Psychological Review 93, 283–321.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. Journal of Memory and Language, 47, 292–314

Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. Language and Speech, 31, 307–319. Language, Cognition and Neuroscience 13

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

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. Language, 80, 748–775. doi:10.1353/lan.2004.0185

Gillespie, M. (2011). Agreement computation in sentence pro- duction conceptual and temporal factors (Unpublished doctoral dissertation). Northeastern University, Boston, MA.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics III: Speech acts (pp. 41-58). New York: Academic Press.

Gussenhoven, C. (1983). A semantic analysis of the nuclear tones of English. Bloomington, IN: Indiana University Linguistics Club.

Halliday, M. A. K. (1967). Intonation and grammar in British English. The Hague: Mouton.

Heller D., Arnold, J. E., Klein N. & Tanenhaus M. K. (2014 online). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech*. doi: 10.1177/0023830914528107

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search*. Journal of Memory and Language*, 58, 541-573. doi:10.1016/j.jml.2007.06.013

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

Kahn, J. & Arnold, J.E. (2012). A Processing-Centered Look at the Contribution of Givenness to Durational Reduction. Journal of Memory and Language, 67, 311–325.

Kahn, J. & Arnold, J.E. (2015). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. Language, Cognition, and Neuroscience.DOI:10.1080/01690965.2013.848989

Ladd, R. (1996). Intonational phonology. Cambridge: Cambridge University Press.

Levelt, W. J. M. (1989). Speaking: From intention to articulation.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech, 6(3), 172–187.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), Speech production and speech modeling (pp. 403– 439). Dordrecht: Kluwer Academic.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication, Proceedings of *the National Academy of Sciences*, vol. 108, no. 9, 2011, pp. 3526.

Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P.R. Cohen, J. Morgan, and M.E. Pollack (Eds.), Intentions in Communication, (pp. 271-311), Cambridge, MA: MIT Press.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole, (Ed.), Radical pragmatics (pp. 223–256). New York, NY: Academic Press.

Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. Natural Language Semantics, 7, 141–177. doi:10.1023/A:1008370902407

Selkirk, E. O. (1996). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), The handbook of phonological theory (pp. 138–151). Oxford: Blackwell.

Seyfarth, R.M. & Cheney, D.L. (1992). Meaning and mind in monkeys. Scientific American, **267**, 122-129.

Seyfarth, R. M. & Cheney, D. L. (this volume). The social origins of language.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing “given” information: Effects of persistence of grammatical function and surface position. Language and Speech, 37, 125–145.

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In L. Frazier & E. Gibson (Eds.), *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor.*New York: Springer International*.*

Watson, D.G., Arnold, J.E., & Tanenhaus, M.K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. Cognition, 106, 156-1557.Chafe, 1976

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. Harvard Studies in Classical Philology, 40, 1-95.

Zerkle, S., Rosa, E. C., & Arnold, J. E. (under review). Thematic role predictability and planning affect word duration. UNC Chapel Hill

1. While some scholars argue that disfluency is produced intentionally as a signal that the speaker is having trouble (Clark & Fox Tree, 2002), there is no theory in which disfluency is a signal of information status per se. [↑](#footnote-ref-1)