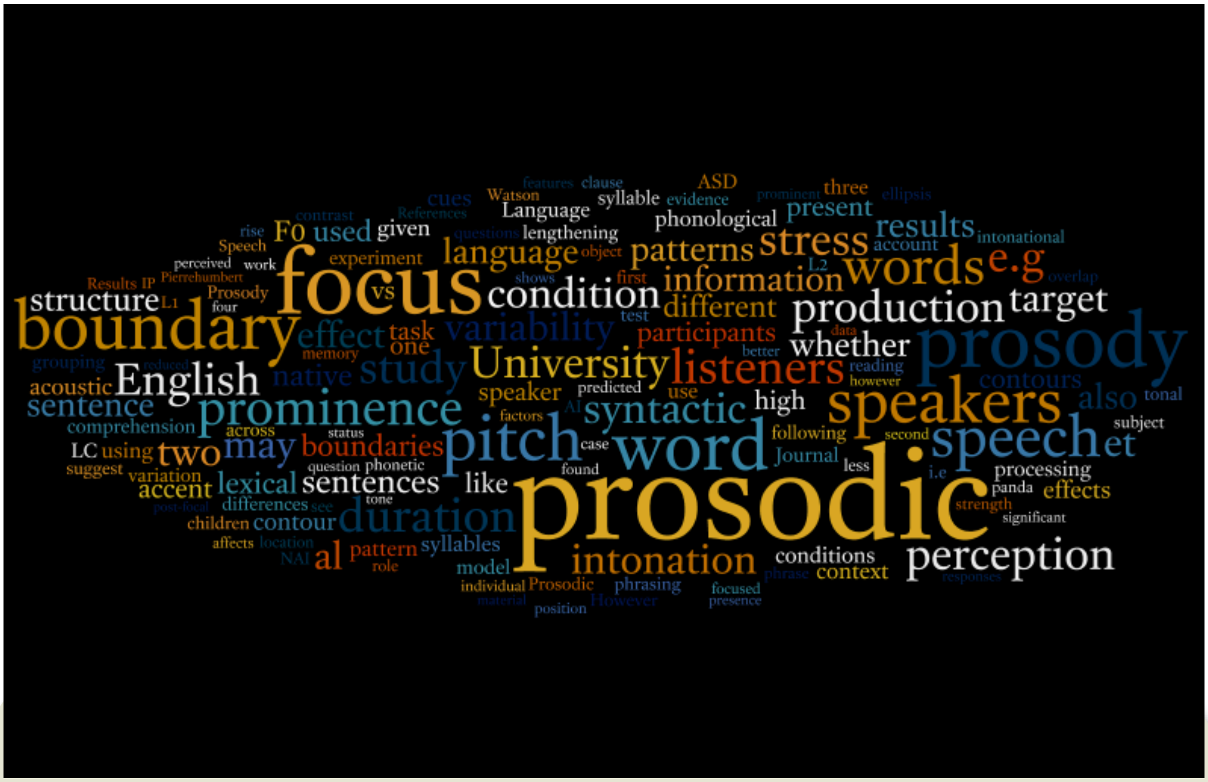

features clause ASD
Watson syllable evidence prominent three ellipsis
cues Language phonological present
contrast References
rise F0 used given questions lengthening results intonational
Speech experiment language object patterns stress account
perceived work focus vs first information words e.g
Results IP Pierrehumbert L2 shows overlap
Prosody condition different production target
structure L1 four
boundary effect task test participants whether prosody
grouping reduced memory one variability data
acoustic English native study University listeners better reading
sentence speaker predicted use however contours also tonal
comprehension prominence factors AI syntactic high speakers subject
LC using two across status pitch case word following second speech et
suggest may boundaries found Journal less processing
variation question phonetic i.e
accent lexical sentences like panda effects
post-focal differences see tone prosodic strength
children contour duration significant
affects location
NAI al pattern syllables intonation conditions perception
role model phrase context responses
individual Prosodic phrasing focused
material position However presence

Experimental and Theoretical Advances in Prosody (ETAP-3)

University of Illinois at Urbana-Champaign

May 28-30, 2015

ACKNOWLEDGEMENTS

Organizing committee

Chigusa Kurumada, Michael Wagner, Duane Watson

Sponsors

We would like to thank the following sponsors for their generous support:

Departments of Linguistics and Psychology
University of Illinois at Urbana-Champaign
http://www.psychology.illinois.edu/
http://www.linguistics.illinois.edu/

College of Liberal Arts and Sciences
University of Illinois at Urbana-Champaign
http://www.las.illinois.edu/

The National Science Foundation

Reviewers

We would also like to thank the following individuals for volunteering as Reviewers:

Jason Bishop, Mara Breen, Meredith Brown, Andrés Buxó-Lugo, Amber Camp, Katy Carlson, Yiya Chen, Lauren Clemens, Mariapaola D’Imperio, Laura Dilley, Emily Elfner, Scott Fraundorf, James German, Maureen Gillespie, Carlos Gussenhoven, Julia Hirschberg, Yuki Hirose, Shin Ishihara, Kiwako Ito, Sun-Ah Jun, Gerrit Kentner, Tuan Lam, Constantine Lignos, Iris Ouyang, Melissa Redford, Amy Schafer, Stefanie Shattuck-Hufnagel, Shari Speer, Jill Thorson, Joseph Tyler, Kristine Yu

PROGRAM

* invited speakers in bold

Day 1 – Thursday May 28

8:30-10:00	Breakfast
10:00-12:00	**Chigusa Kurumada** (University of Rochester)……………...5
	Processing and interpreting the variable prosodic signal
	Jennifer Arnold and Elise Rosa……………………………………………………..10
	Mechanisms of prosodic variation: Selectional vs. fluency accounts
	Amelia Kimball, Jennifer Cole, Gary Dell, and Stefanie Shattuck-Hufnagel…………...11
	Is phonetic variation represented in memory for pitch accents?
	Kristine Yu and Ed Stabler……………………………………………………………12
	A parsing model for crowding, speech rate and syntax of tone
12:00-2:00	Lunch / Poster Session
2:00-3:30	**Alice Turk** (University of Edinburgh)…………………………………………………5
	Patterns of variability as evidence for extrinsic timing in speech motor control
	Sandra Zerkle, Elise Rosa, and Jennifer Arnold………………………………………13
	Thematic-role predictability and planning affect word duration
	José Hualde, Jennifer Cole, Tim Mahrt, and Christopher Eager………………………...14
	Early high and prominence perception
3:30-4:30	Coffee
4:30-5:30	**Naomi Feldman** (University of Maryland)……………………………………………….6
	Testing low-level speech features using speech corpora
	Aoju Chen…………………………………………………………………………......16
	Children differ in prosodic realisation of focus: How and why?
7:00-	Dinner at Big Grove (registration required)

Day 2 - Friday May 29

8:30-10:00	Breakfast
10:00-12:00	**Tyler Kendall** (University of Oregon)…………………………………………………....6
	Corpus-based approaches to variation in speech rate and pause
	Natalie Boll-Avetisyan, Anjali Bhatara, and Barbara Höhle……………………………17
	Effects of dyslexia and musicality on the perception of rhythmic structure
	Mara Breen and Johanna Kneifel……………………………………………………...18
	Rhythmic context affects on-line ambiguity resolution in silent reading
	Loretta Yiu and Duane Watson………………………………………………………..19
	Phonological overlap affects word duration in discourse contexts
12:00-2:00	Lunch / Poster Session

2:00-3:30 Alejna Brugos and Jonathan Barnes……..20

Cross-phrase tonal patterns cue boundary strength (variably)

Sarah Bibyk, Christine Gunlogson, and Michael Tanenhaus……22

SET it up: Speaker knowledge and intonational contours in a natural production task

Kiwako Ito, Rory Turnbull, and Shari Speer……24

Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses

3:30-4:00 Coffee

4:00-5:30 **Morgan Sonderegger** (McGill University)……7

The role of prosodic variability in explaining segmental variability: Two corpus studies

Nazik Dinctopal-Deniz and Janet Fodor……25

Prosodic disambiguation in Turkish: Evidence from phoneme restoration

Emily Elfner……...26

Prosodic juncture strength and syntactic constituency in Connemara Irish

Day 3 - Saturday May 30

10:00-10:15 Coffee

10:15-12:45 **Mark Liberman** (University of Pennsylvania) …....8

A cross-linguistic study of prosodic focus

David Lutz and Sam Tilsen……...28

Semantic expectation affects prosodic prominence perception

Andrés Buxó-Lugo and Duane Watson……29

Top-down processing of intonational boundaries

Aron Hirsch and Michael Wagner……30

Syntactic constraints on the variability of prosodic phrasing and parenthetical placement

1:00-2:00 Lunch

2:00-3:30 Panel Discussion

3:30-5:00 **Jennifer Cole, Timothy Mahrt, and Joseph Roy** (University of Illinois)......................8

New methods of crowd-sourcing for prosodic annotation: Inter-annotator agreement, individual differences, and sources of variation

POSTER SCHEDULE

Thursday May 28 12-2pm

Poster #	Poster Title	Page #
1	Jason Bishop, Adam Chong, and Sun-Ah Jun *Prosodic phrasing and individual differences in relative clause attachment*	34
2	Amelia Kimball, Loretta Yiu, and Duane Watson *Metrical regularity of surrounding context affects word recall*	35
3	Elizabeth Pratt and Eva Fernandez *Implicit prosody and task-load interactions in native and late bilingual speakers of English*	36
4	Amanda Rysling, Charles Clifton, and Lyn Frazier *Ellipsis incurs a penalty in parentheticals*	38
5	Fabian Schubö *Prosodic phrasing in German narrow focus constructions*	39
6	Aaron Albin *Typologizing native language influence on second language intonation production: Three transfer phenomena in Japanese EFL learners*	40
7	Anca Chereches Some *clustering with k-means: Acoustics of a prosodic contrast in a corpus of spontaneous speech*	41
8	Jonathan Barnes, Nanette Veilleux, Alejna Brugos, and Stefanie Shattuck-Hufnagel *Interpreting patterns of variability in the realization of English intonation contours*	42
9	Vsevolod Kapatsinski and Paul Olejarczuk *Perceptual learning of intonation contour categories: The role of utterance duration*	44
10	Jonathan Howell *Double focus of adjacent words in yes/no questions*	45

Friday May 29 12-2pm

Poster #	Poster Title	Page #
1	Shinobu Mizuguchi, Jennifer Cole, Gabor Pinter, Koichi Tateishi, and Tim Mahrt *Natural speech perception by L1 and L2 speakers of English*	46
2	Philippe Martin *Accent phrases duration constraints*	48
3	Joseph Tyler and Ezra Keshet *Prosodic disambiguation of conditional vs. logical conjunction*	49
4	Aisha Bhutto and Maureen Gillespie *Perceived prosodic variability predicts gesture rates in spontaneous speech*	50
5	Donna Erickson, Albert Rilliard, Takaaki Shochi, and João Moraes *Prosodic strategies of L1 and L2 speakers for attitudinal expressivity in USA English*	51
6	Kathryn Weatherford and Jennifer Arnold *Audience design effects on prosody*	52
7	Ayako Kondo, Takayuki Konishi, Ken'ya Nishikawa, Hidetoshi Takahashi, Yoko Kamio, and Reiko Mazuka *Variability as a key feature of Autism spectrum disorders prosody*	53
8	Tuuli Morrill *Native language effects on phrasal intonation: Application of smoothing spline analysis of variance (SS ANOVA) to multi-syllabic utterances*	54
9	Argyro Katsika *Lexical stress and the scope of boundary lengthening*	55
10	Mariapaola D'Imperio and James German *Dialect imitation across typologically distinct prosodic systems*	56

INVITED SPEAKERS

-----Thursday, 10-10:30am-----

Processing and interpreting the variable prosodic signal

Chigusa Kurumada (University of Rochester)

Variability is ubiquitous in the prosodic signal (Liberman & Pierrehumbert, 1984). Some of the variation is due to random factors, such as noise and errors in production. However, much of the variability stems from systematic factors such as age and gender of the speaker, dialects/accents, speech conditions and contexts (e.g., Ladd 2008). The resulting lack of invariance between the signal and phonological representations (e.g., pitch accents) has presented challenge to many approaches (see Cole, 2015 for an overview). In addition, it further complicates the problem that listeners' interpretive domains tend to overlap. For instance, while listeners typically associate H* with new information in discourse, it can also be compatible with a contrastive interpretation (Watson, Tanenhaus & Gunlogson, 2008).

In this overview talk, I will consider how listeners may navigate the variability in the prosodic signal to derive coherent interpretations. Specifically, I will outline a perspective in which listeners flexibly adapt their interpretation of the bottom-up acoustic signal based on their recent experiences and top-down inferences based on contextual information. The goal of this talk is to discuss existing evidence of flexible interpretations of prosodic information (e.g., Brown, Dilley & Tanenhaus, 2014; Kurumada et al., 2012, 2014; Ito, Arai, & Hirose, 2015) in light of recent theoretical approaches to variability and adaptation in speech perception research. In so doing I propose some ways in which examining prosodic variability can advance our understanding of the mechanisms underlying prosodic processing.

-----Thursday, 2-2:30pm-----

Patterns of variability as evidence for extrinsic timing in speech motor control

Alice Turk in collaboration with Stefanie Shattuck-Hufnagel (University of Edinburgh)

Articulatory Phonology/Task Dynamics (AP/TD, Browman & Goldstein 1985, Saltzman & Munhall 1989 et seq.) is the model of speech production that currently provides the most comprehensive account of speech timing phenomena. Timing control in this model is intrinsic, that is, time is an inherent part of phonological representation. Surface timing patterns emerge from properties of the system and do not need to be specified, tracked, or modified during an utterance. However, several lines of behavioral evidence challenge intrinsic timing as implemented in Articulatory Phonology/Task Dynamics, and support the view that timing control in speech production is extrinsic. This evidence comes from patterns of invariance and variability, and includes 1) evidence that phonological representations are symbolic, 2) temporal evidence for separate control of movement onsets and offsets, difficult to implement in mass-spring systems such as AP/TD 3) increasing variability with increases in interval duration, as predicted by a "noisy timekeeper" model, 4) language-specific constraints on surface timing in a quantity language, not predicted in intrinsic timing systems which do not allow surface timing specifications, and 5) motor equivalence of strategies for producing surface duration patterns, again difficult to explain in intrinsic timing systems with no reference to surface time. These lines of evidence motivate the consideration of alternative models. A brief sketch of an alternative extrinsic timing model of speech motor control will be presented.

This approach involves three distinct phases: 1) Phonological planning, 2) Phonetic planning, and 3) Phonetic implementation. The phonological planning stage involves structuring symbolic segmental representations into a hierarchy of prosodic constituents and prominences. We assume that phonetic planning involves balancing task requirements and movement costs to yield (near-) optimal parameter values (cf. Optimal Control Theory approaches, e.g. Todorov & Jordan 2002) for use in the third, Phonetic implementation stage. Task requirements include things like being accurate, spreading information evenly throughout the signal via an appropriate prosodic structure, and not taking too long. Movement costs include things like the spatial inaccuracy cost of moving fast, and energy expenditure. The phonetic planning stage involves planning a sequence of goal states (e.g. spectral properties that can serve as cues to planned distinctive features, cf. Guenther 1995), the timing between goal states, the articulators that create the goal states, spatial goals of their movements, and movement timing characteristics, including the timing of movement onsets and movement time course characteristics (Lee 1998). In this approach, articulatory overlap results from movement goals which follow each other in (relatively) rapid succession. As speech unfolds, we assume that speakers continuously track their movements to reach their targets with desired spatial and temporal accuracy (cf. Bullock & Grossberg 1988).

-----Thursday, 4:30-5pm-----

Testing low-level speech features using speech corpora
Naomi Feldman (University of Maryland)

Listeners generalize across utterances from different speakers, dialects, listening conditions, and contexts. I show how a cognitive model that operates over speech corpora can be used to evaluate hypotheses about the dimensions of speech that guide these generalizations. Simulations show that representations that are normalized across speakers predict human discrimination data better than unnormalized representations, mirroring previous findings. The model also reveals differences across normalization methods in how well each predicts human data. These results indicate that cognitive modeling can be used to quantitatively evaluate different representations of speech, yielding consistent and interpretable results.

-----Friday, 10-10:30am-----

Corpus-based approaches to variation in speech rate and pause
Tyler Kendall (University of Oregon)

This talk considers how the examination of prosodic features through corpus-based analyses can inform theoretical and experimental approaches to prosody. Following from work reported in Kendall (2013), I discuss variation in speech timing through the analysis of conversational speech from over 150 American English speakers, mobilizing a diverse set of sociolinguistic recordings housed in the Sociolinguistic Archive and Analysis Project (SLAAP; Kendall 2007). I begin by providing a brief overview of SLAAP and its data and tools and then survey a number of corpus-based inquiries into variation in articulation rate and silent pause duration in the conversational recordings made possible by spoken language corpora with time-aligned transcription. These inquiries demonstrate that articulation rate variation is highly patterned by social and interactional factors (e.g. speaker age, sex, ethnicity, regional origin, sex of the interlocutor) as well as utterance-internal factors (utterance length), at both an utterance- and speaker-level. Silent pause durations, however, are much less patterned at the token- or speaker-level, with only limited effects for social factors (e.g. regional origin, sex) or internal factors (articulation rates, pause frequency). These findings are inline with accounts in the literature that associate pausing with cognitive factors in language production (Goldman-Eisler 1968, Rochester 1973, Kowal and O'Connell 1980, Levelt 1989, Krivokapic 2007, Redford 2013) and lend support for a view that articulation rates are less interrelated with cognitive factors.

To consider the utility of corpus-based analyses for theoretical and experimental approaches to speech timing further, I then focus on a relatively unexamined question in the research on pausing: *how long is a pause*? This question has important methodological ramifications for studies of pausing (e.g. what thresholds are to be used for determining which silences in speech should be accounted for in an analysis of pause? Cf. Campione and Véronis 2002), as well as larger implications for theories that attempt to relate pausing to speech production. In particular, I present a modeling experiment using the almost 30,000 silent pause measurements of the corpus study to assess the distribution of pause durations in conversational speech. This yields the finding that silent pauses shorter than ~0.5 second bear a different relationship to predictive factors than longer pauses, suggesting that silent pauses in conversational speech occur in (at least) two distributions and that each of these distributions is potentially influenced by different factors. In other words, very short pauses and less short pauses may be manifested by speakers for different reasons.

Throughout the talk, I highlight questions that are raised by these corpus-based studies for future work on these speech timing features and prosody more generally.

-----Friday, 4-4:30pm-----

The role of prosodic variability in explaining segmental variability: Two corpus studies

Morgan Sonderegger (McGill University)

Research on speech increasingly seeks to understand the massive variability present in the signal, within and outside the laboratory. This conference's focus on prosodic variability is paralleled by much recent work on segmental variability, including in the areas of sound change (diachronic variability in how sounds are produced in a speech community) and phonological/phonetic variation (synchronic variability in `` ``). A key question driving research in each area is, how can we explain patterns of variability in segmental realization across speakers and over time? This talk addresses this question for two cases of sound change and phonological variation observed in speech corpora, where prosodic variability is crucial for explaining segmental variability.

I first discuss a case of sound change in Seoul Korean where segmental variation is turning into prosodic variation, a.k.a. tonogenesis. The difference between "aspirated" and "lax" stops, which was historically realized as a VOT difference using ([pul] vs. [p^{h}ul]), is increasingly realized as an f0 difference on the following vowel ([pùl] vs. [púl]). Previous work has documented that this change is spreading through the community (Kang, 2014; Silva, 2006), but little is known about how and why the tradeoff between segmental and prosodic cues unfolds over time. We use a corpus of read speech by 120 Seoul Korean speakers aged 20s-60s, to examine how the change has unfolded across the lexicon, focusing on the effect of word frequency. We find that VOT reduction and f0 enhancement are greatest in higher-frequency words, suggesting that this change is driven by hypo-articulation (which affects VOT). We also find a tight coupling between VOT reduction and f0 enhancement over time, suggesting that f0 enhancement is an adaptive response to the erosion of VOT as a cue, in order to maintain the contrast. The overall picture is of prosodic variability being "recruited" by speakers to compensate for increasingly noisy segmental variability.

I next discuss a study examining the relationship between prosodic boundaries and segmental variation through the lens of deletion of word-final coronal stops in consonant clusters in English (CSD; e.g. fast realized as [fæs] or [fæst]). CSD rate has been shown to be affected by many factors, most importantly the phonological context (e.g. [fæs] more likely in fast car vs. fast one). Previous work on variable segmental realization, including CSD, often operationalizes the role of prosodic boundaries as a binary "following pause" (between the coronal stop and the next word), which is treated as one type of context, on par with a following consonant or vowel. We examine CSD rates in a corpus of spontaneous British English speech from 20 speakers, instead treating the length of a pause as a quantitative proxy for boundary strength, and test the hypothesis that it does not only by itself affect deletion rate, but also modulates the effect of the following segment, as predicted if phonological/phonetic processes are constrained by the locality of production planning (Wagner, 2012). The results support this hypothesis, and illustrate a promising new direction for understanding the role of prosody in determining when and how segments are produced variably.

-----Saturday, 10:15-10:45am-----

A cross-linguistic study of prosodic focus

Mark Liberman (University of Pennsylvania)

We examine the production and perception of (contrastive) prosodic focus, using a paradigm based on digit strings, in which the same material and discourse contexts can be used in different languages. We find a striking difference between languages like English and Mandarin Chinese, where prosodic focus is clearly marked in production and accurately recognized in perception (95-97%), and languages like Korean and Suzhou Wu, where prosodic focus is neither clearly marked in production nor accurately recognized in perception (45-55%). We also present comparable data for Japanese and French.

This typological dimension is apparently not predicted by standard typological distinctions such as the presence or absence of lexical tone, the presence or absence of lexical stress, dynamic vs. melodic accent, etc. This may reflect a fundamental typological difference in the prosodic marking of focus, or it may be a consequence of other (perhaps unrecognized or misunderstood) aspects of prosodic typology.

Reference:

Lee, Yong-cheol, Bei Wang, Sisi Chen, Martine Adda-Decker, Angélique Amelot, Satoshi Nambu, and Mark Liberman, "A Cross-linguistic Study of Prosodic Focus", IEEE ICASSP 2015.

-----Saturday, 3:30-5pm-----

New methods of crowd-sourcing for prosodic annotation: Inter-annotator agreement, individual differences, and sources of variation

Jennifer Cole, Timothy Mahrt, & Joseph Roy (University of Illinois at Urbana-Champaign)

This talk presents methods for crowd-sourcing prosodic annotation from untrained listeners using Rapid Prosody Transcription (RPT), and demonstrates web-based tools for deploying annotation tasks in the lab and on the internet. We show how agreement analyses with Kappa statistics can be used to determine the optimal number of annotators needed to maximize the reliability of crowd-sourced annotations and to assess how inter-annotator agreement varies in relation to the size of the annotated database and annotator cohort (lab vs. internet, within- or across-dialect). Regression modeling with generalized linear mixed models and generalized additive models is introduced to explore individual and cohort-level differences in the factors that predict a word’s binary prosodic labels. With RPT carried out in real-time, these models are also able to capture the effect of time-lag as a factor in prosody perception. A comparison of crowd-sourced prosodic annotations with ToBI annotations from trained annotators validates the use of coarse-grained prosodic labels as approximations of fine-grained phonological annotation. We end with a brief mention of recent studies using RPT for prosodic analysis with under-resourced languages and languages for which a phonological prosodic analysis has not yet been undertaken.

Oral Sessions

Mechanisms of prosodic variation: Selectional vs. fluency accounts

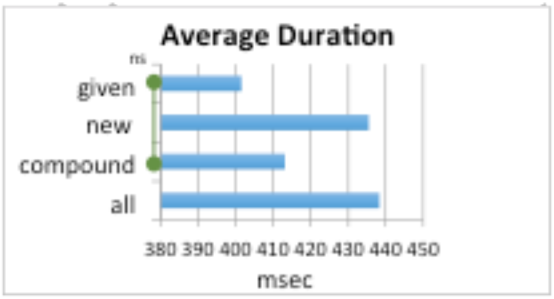
Jennifer E. Arnold & Elise C. Rosa (UNC Chapel Hill)

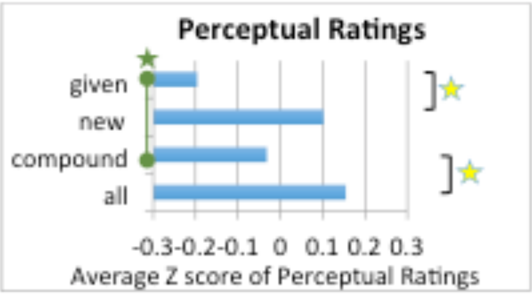
Speakers mark information status prosodically. For example, in "The panda blinks", the pronunciation of "panda" is likely to be reduced (faster and less articulated) when it is "given" (mentioned) than when it is new. The standard explanation draws on the topical/in-focus status of given information vs. the out-of-focus status of new information (e.g., Breen et al., 2010; *LCP*). This view suggests a selectional mechanism: prosodic form is selected to communicate information status. An alternate explanation is fluency: subsequent productions of "the panda" may be shorter because the words are lexically and phonologically available, and thus easier to produce (Arnold & Watson, 2014, *LCN*, Kahn & Arnold, 2012, *JML).* We test these accounts in a production experiment. Participants described images of animals performing pairs of actions (spin, expand, blink, shrink), in four conditions. The target is underlined.

COMPARISON GROUP 1: SINGLE ANIMALS
1) Given: The panda spins. The panda blinks.
2) New: The frog spins. The panda blinks.

COMPARISON GROUP 2: MULTIPLE ANIMALS
3) Compound: The panda and the frog spin. The panda blinks.
4) All: All the animals spin. The panda blinks.

Both accounts predict that the target "panda" should be acoustically reduced in the given vs. new condition. The condition of interest is the compound condition. Informationally, the compound condition patterns with the new condition, in that the target is not in focus. The compound should thus lead to no reduction in comparison with the multiple-animal comparison case, the "All" condition. By contrast, the fluency account predicts that both the given and compound conditions should be reduced due to lexical facilitation of the target word. We analyzed 2 measures: 1) log duration, and 2) perceptual ratings (the average z-score of 6 ratings by trained RAs on a scale of 1 to 3).





Robust effects of fluency emerged on the target duration, which revealed a main effect of lexical repetition (given/compound shorter than new/all), and no interaction with number of animals – i.e., given and compound conditions were equally shortened. The perceptual ratings also reflected fluency, in that both given and compound were less prominent than new/all. At the same time, the ratings suggested that the compound condition was somewhat less reduced than the given condition: lexical repetition interacted with number of animals, and a planned contrast revealed that the given condition was significantly less prominent than the compound condition. This is consistent with the idea that information status selects for more prominent forms in this compound condition, which is not in focus – despite lexical facilitation. These findings suggest that both fluency and selectional mechanisms contribute to prosodic variation, each affecting different measures.

Is phonetic variation represented in memory for pitch accents?

Amelia Kimball[1], Jennifer Cole[1], Gary Dell[1], & Stefanie Shattuck-Hufnagel[2]
[1] University of Illinois at Urbana-Champaign, [2] MIT

Phonological accounts of prosody postulate that listeners map variable instances of prosody to categorical features. Other research maintains that listeners remember subcategorical phonetic detail [1,2]. Our study probes memory to investigate the reality of categorical encoding for prosody-- when listeners hear a pitch accent, what do they remember? Two types of prosodic variation are tested: phonological variation (presence vs. absence of a pitch accent), and variation in phonetic cues to pitch accent (F0 peak, duration). We report results from six experiments that test memory for pitch accent vs. cues.

Stimuli were nouns excised from natural productions of sentences of American English. Twelve words were recorded in accented and unaccented forms. Each accented word was resynthesized to create a large phonetic change (+/- 20 Hz F0 peak height or +/- 10 percent change in duration) without changing phonological accent status.

Experiments were conducted online using Amazon Mechanical Turk. In experiments 1-3 participants hear two words and report if they are exactly the same or different (an AX task). Experiments 4-6 use the same stimuli but add a delay and interference to increase memory demands. Listeners hear four different words (exposure), then a tone, and then another presentation of a word from the exposure phase (test). They report whether the test word is exactly the same as the exposure version.

Listeners are well above chance at discriminating all three contrasts in the AX task. Furthermore, performance does not differ significantly for phonological and phonetic differences. When comparing performance in the AX task to the delayed task, listeners do not differ significantly in their recognition accuracy for accent (76% AX vs. 83% delay, N=30), but are statistically significantly worse at recognizing both pitch (75% AX vs. 54% delay, N=30) and duration (85% AX vs. 67% delay, N=30). This suggests that after a delay and interference phonetic details are less accessible than phonological accent status. This evidence is consistent with the hypothesis that listeners encode detailed instances of pitch accents, but that details fade in memory more quickly than categorical distinctions.

Group effects hold when analyzed with a mixed effect logit model with random slopes and intercepts to account for individual variability. However, examining individual performance in the AX task shows that listeners' memory for prosodic features is variable. The standard deviation of scores in the AX pitch task was significantly higher than the standard deviation of scores in the AX duration task (F(29,29) =2.7492, p<.01) or AX accent task(F(29,29) =3.1501, p<.01), meaning performance varied more from listener to listener in the pitch task than in the duration or accent task. This holds despite the fact that these same listeners were excellent at discriminating a pure tone difference of the same magnitude pitch (mean=91%, s.d. =.133%).

Taken together, our results suggest that (1) memory may be a useful tool for investigating the representation of prosody (2) listeners encode both categorical distinctions and phonetic detail, but categorical distinctions are more accessible at retrieval and (3) listeners may vary in the degree to which they remember prosodic detail.

[1] Goldinger, S.D. (1996) Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166-1183.
[2] Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology, 70,* 1-30.

A parsing model for crowding, speech rate and syntax of tone

Kristine M. Yu (UMass Amherst) & Edward Stabler (UCLA)

Preliminary models of intonational phonology of languages often abstract away from the pervasive variability in the mapping between the speech signal and grammatical prosodic structures. But of course, any language comprehension model must deal with this variability. As a first step towards such a model, this paper takes as a case study the allophonic variability conditioned by tonal crowding and speech rate in Samoan and presents a computationally implemented parsing model that can handle the interaction of crowding, speech rate and syntax in tonal realization. The model is 'traditional' in that it takes the phonetic signal and some representation of context as input and produces (possibly weighted) structures as output (Halle and Stevens, 1962; Fodor, Bever, and Garrett, 1974). It is sometimes suggested that it is difficult or impossible for models that are 'traditional' in this sense to account for this kind of fine-grained phonetic variability – see for example Cutler (2012, p.228) and Farmer, Brown, and Tanenhaus (2013); Brown, Dilley, and Tanenhaus (2012) – but this paper shows how a traditional performance model offers an integrated account of the interactions of symbolic and phonetic factors.

Samoan is an ergative Austronesian language that has been described as having an unmarked absolutive case (Chung, 1978; Bittner and Hale, 1996), but, following Legate (2008), Collins (2014) argues that what appeared to be absolutive is actually nominative and accusative, and Yu (2009, 2014) argues that the absolutive in Samoan is actually marked, but by a high tone rather than by affixes or other segmental material. Thus, Samoan has a tonal case morpheme, i.e. syntax directly conditions this aspect of the language's intonation. The prosodic case marker can be obscured by various kinds of conditioned variability. Depending on its surrounding segmental material and real-time separation from other tones, the high tone can be 'crowded' and obscured (Bruce, 1977; Pierrehumbert, 1980; Arvaniti, Ladd, and Mennen, 2006; Gordon, 2014), even though it is clearly present in other contexts.

This paper shows how a simple, 'traditional' parsing model (Stabler, 2013) can be extended to accommodate this kind of variability. Candidate segmentations of the signal are parsed top-down and incrementally to yield (partial) derivations of each potential segmentation, each of which is used to concurrently and incrementally compute prosodic, phonetic and semantic representations (Shieber, 2014). One of the few explicit descriptions of algorithms for mapping between fundamental frequency (f0) contours and intonational representations, that of Pierrehumbert and Beckman (1988), is adapted to model tonal crowding in Samoan, predicting a tonal realization that is compared to the segmented signal and used to weight alternative (partial) parses. Crucially, the tonal case marking is predicted to have a distinct realization only when crowding and other factors allow. Like Pierrehumbert and Beckman (1988), the algorithm constructs a sketch of the f0 contour where tones are instantiated as target f0 levels. However, while Pierrehumbert and Beckman (1988) hard-codes speaker-dependent parameters that determine these target levels, the parameters here are also a function of the phonetic context and set by cross-validation. Moreover, variability in the 'sag' between high tonal targets due to tonal crowding cannot be captured by linear interpolation (as in Pierrehumbert and Beckman (1988)); instead, transitions between targets are computed using quadratic functions (Pierrehumbert, 1981).

Arvaniti, A., et al. 2006. Phonetic effects of focus and 'tonal crowding' in intonation. *Speech Comm.*, 48:667–696. Bittner, M. and K. Hale. 1996. The structural determination of case and agreement. *LI*, 27:1–68. Brown, M. et al. 2012. Real-time expectations based on context speech rate can cause words to appear or disappear. *Procs 34th Ann. Conf. Cog. Sci. Soc.*, pages 1374–1379. Bruce, G. 1977. *Swedish word accents in sentence perspective*. CWK Gleerup, Lund. Chung, S. 1978. *Case Marking and Grammatical Relations in Polynesian.* U. Texas Press, Austin, TX. Collins, J. N. 2014. The distribution of unmarked cases in Samoan. *AFLA* 12, pages 93–110. Cutler, A. 2012. *Native Listening*. MIT Press, Cambridge, MA. Farmer, T.A. et al. 2013. Prediction, explanation, and the role of generative models in language processing. *BBS*, 36:211–212. Fodor, J. A. et al. 1974. *The Psychology of Language*. McGraw-Hill, NY. Gordon, M. 2014. Disentangling stress and pitch accent. In H. van der Hulst, editor, *Word Stress*. Cambridge U. Press, NY, pages 83–118. Halle, M. and K. N. Stevens. 1962. Speech recognition: A model and a program for research. *Trans. of the Prof. Group on Info. Theory*, IT-8:155–159. Legate, J. A. 2008. Morphological and abstract case. *LI*, 29(1):55–101. Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, Harvard U. Pierrehumbert, J. 1981. Synthesizing intonation. *JASA*, 70:985–995. Pierrehumbert, J. and M. Beckman. 1988. *Japanese Tone Structure*. The MIT Press. Shieber, S. M. 2014. Bimorphisms and synchronous grammars. *J. Lang. Modelling*, 2(1):51–104. Stabler, E. P. 2013. Two models of minimalist, incremental syntactic analysis. *TiCS*, 5(3):611–633. Yu, K. M. 2009. The sound of ergativity: Morphosyntax-prosody mapping in Samoan. *NELS* 39. Yu, K. M. 2014. Tonal marking of absolutive case in Samoan. ETI3.

Thematic-role predictability and planning affect word duration

Sandra A. Zerkle, Elise C. Rosa, & Jennifer E. Arnold (UNC Chapel Hill); szerkle@live.unc.edu

It is well established that word duration is influenced by the predictability of a word, or its probability within context [1, 2, 3]. However, little is known about the effects of thematic role predictability on word duration (but see [4]), and nothing about the thematic roles of transfer verbs, as in (1). In sentences like these, the goal (Sir Barnes) is more likely to be mentioned again [5].

1a) Lady Mannerly gave a painting to Sir Barnes.

1b) Sir Barnes received a painting from Lady Mannerly.

Our first goal was to test whether the predictability of goals leads to acoustic reduction, using a naturalistic production task. Our second goal was to test the hypothesis that predictability affects acoustic reduction by modulating the difficulty and timecourse of speech planning [6].

In two experiments, participants were told they were tabloid photographers, and were asked to help a detective to help solve a crime by describing pairs of "their pictures". The detective described the first picture, and the participant described the second. For critical items, the first panel illustrated a transfer event as in (1), and the second illustrated an event with either the source or goal, e.g. "Sir Barnes threw it in the closet". In Exp1, the participant only viewed the pictures on a computer screen. In Exp2, the detective enacted the story on an interactive magnet boards with movable characters. This added an audience design factor and made the conversation more engaging. It also delayed the point when participants could begin planning the utterance, which is likely to have increased the degree of incremental planning. This analysis is limited to responses that a) used a description for reference (e.g., *Sir Barnes)*, and b) referred to the nonsubject character (since references to the subject were usually pronominalized). Our dependent measures were a) log latency of utterance initiation, and b) log of target duration. A separate group of 20 participants viewed the pictures and rated the likelihood of mentioning the target character, and rated the goal continuations as significantly more likely than source continuations (p=0.0015). These ratings were included as predictors.

In Exp1, the only effect on target duration was one of onset latency (p=0.009): target durations were shorter on trials with shorter latencies. In Exp2, again longer latencies predicted longer durations (p=0.005). In addition, there was a goal by likelihood interaction effect on duration (p=0.002), such that goal durations are shorter than sources when they are more likely.

Together, these findings reveal a robust effect of planning time on referent duration. While predictability (thematic role and likelihood) also led to acoustic reduction in Exp2, this effect was less consistent. This provides strong support for a planning mechanism. By contrast, evidence that predictable referents are reduced is more variable, and possibly dependent on the degree of incremental planning. Nevertheless, thematic roles also affected latency itself – in both experiments, goal trials were initiated faster than source trials. Thus, planning facilitation may mediate the effect of thematic role predictability.

Figure 1

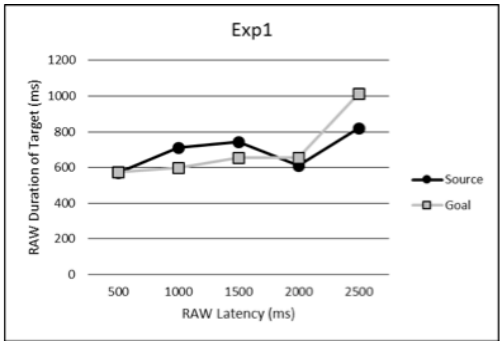
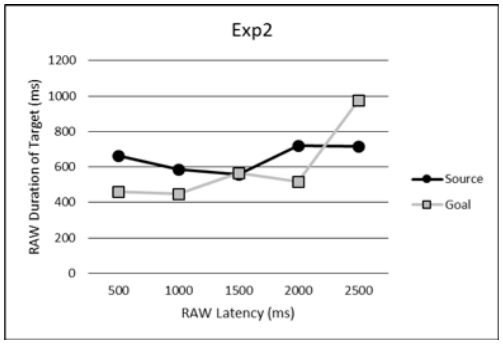


Figure 2



References

1. Bell et al. (2009). Predictability effects on durations of content and function words in conversational English. *JML*. 2. Gahl & Garnsey. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language.* 3. Jurafsky et al. (2000). Probabilistic Relations between words: Evidence from Reduction in Lexical Production. *Typological studies in language*… 4. Kaiser, Li, & Holsinger. (2011). Exploring the Lexical and Acoustic ..… 5. Kehler et al. (2008,). Coherence and Coreference Revisited. *J. Semantics.* 6. Arnold & Watson (2015). *Synthesizing meaning…,* LCN.

Early high and prominence perception

José Hualde, Jennifer Cole, Tim Mahrt, & Christopher Eager (University of Illinois at Urbana-Champaign)

In English, words where the syllable with primary stress is preceded by another syllable with secondary stress (2-1 words, e.g. *èlevátion, òptimístic*) allow a less frequent pronunciation with reversal of prominence (e.g. *èlevátion* → *élevàtion*), in which case their pattern becomes identical to that of 1-2 words (e.g. *élevàtor, súpermàrket*). The perceptual stress reversal phenomenon is produced by the association of a High pitch-accent with the syllable that bears secondary stress in the unmarked pronunciation of the word. We will thus refer to this phenomenon as "early high" (see [1]). Early high in 2-1 words is reported in stress clash contexts ([1], [2]), but is not limited to this context. It is for instance frequent in "news broadcaster style" and it also appears in conversational contexts without stress clash. A question that arises is what motivates the application of early high outside of contexts of stress clash. Here we test the hypothesis that early high affects prominence in a nonlocal fashion.

We have constructed 90 sentences starting with noun phrases such as *The university of Kenya*, where the first word has a lexical 2-1 stress pattern. Each of these phrases was recorded three times, modifying the realization of the first word (Accent Pattern 1 = H* accent on syllable with primary stress, Accent Pattern 2 = early high, and Accent Pattern 3 = unaccented) and the three modifications were spliced with the same continuation of the sentence, for a total of 270 stimuli. For this experiment we chose 30 stimuli, divided in 3 blocks of 10. A total of 30 subjects participated in a Rapid Prosody Transcription task (see [3]), where they were asked to identify those words that they perceived as prominent. Subjects were evenly divided into three groups. Each group heard 10 utterances for each of the three accent patterns.

We have two main results, illustrated here with the example NP *the University of Kenya*: 1. The first content word (*university*) is most often perceived by our listeners as prominent when it bears a H accent on the syllable with primary lexical stress, is slightly less often perceived as prominent when the H is anchored on the syllable with secondary stress, and almost never perceived as prominent when unaccented. 2. Conversely, a following content word in the same NP (*"Kenya"*) is nearly always perceived as prominent when the first content word (*"university"*) is unaccented. The early or late location of the accentual H on *University*, on the other hand, does not affect the perceived prominence of *Kenya* significantly, although there is a trend for a H accent on the primary stress syllable to reduce the perception of prominence on *Kenya* to a greater extent. A possible interpretation is the early high pattern combines the functions of marking the beginning of a phrase and lending prominence to the word bearing it.

Since only the first content word of each sentence was prosodically manipulated, our results offer strong evidence for the relational character of stress at the phrasal level.

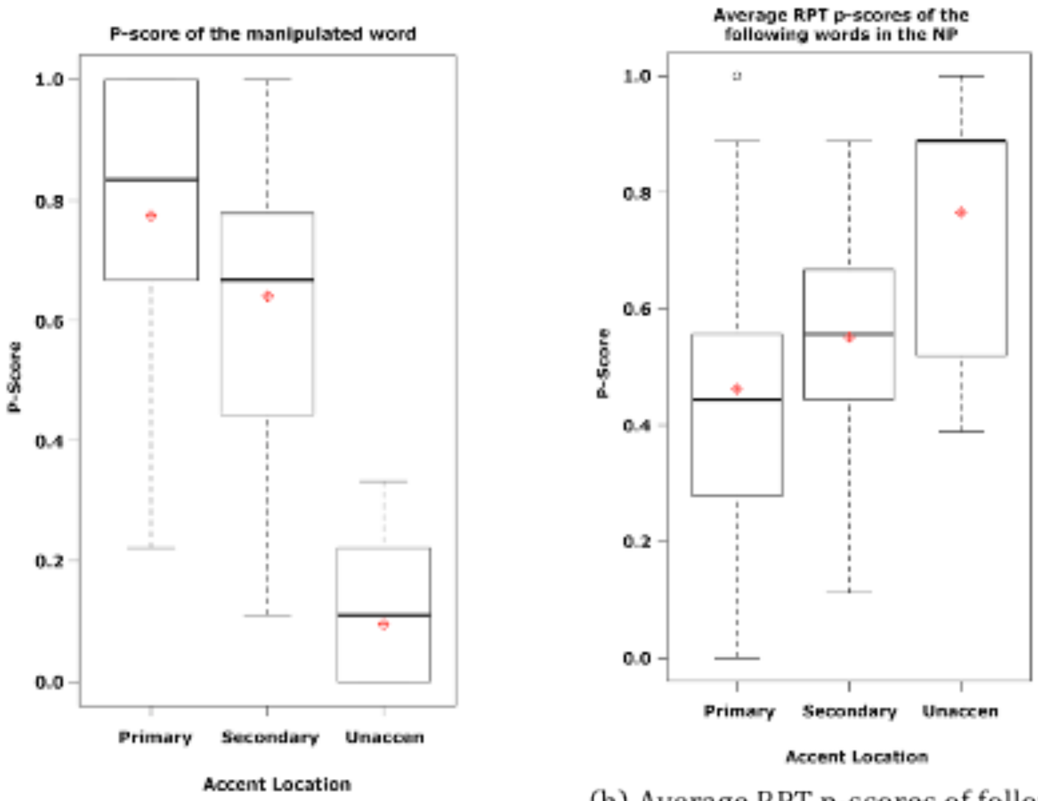


Fig. 1. (a) RPT p-score of the manipulated word (“*University*”).

(b) Average RPT p-scores of following words in the NP by accent pattern of manipulated word ("*Kenya*")

References [1] Shattuck-Hufnagel, S. 1998. “Acoustic-phonetic correlates of stress shift.” Journal of the Acoustical Society of America 84.S1: S98-S98./[2] Beckman, M. & Edwards, J. “Articulatory evidence for differentiating stress categories”. In Keating, P., ed., Papers in Laboratory Phonology III, 7-33. /[3] Cole, J., Mo, Y., & Hasegawa-Johnson, M. 2010. “Signal- based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1:425-452, 2010

Children differ in prosodic realisation of focus: How and why?

Aoju Chen (Utrecht University)

Research on acquisition of prosodic marking of focus (i.e. new information in a sentence) has been primarily concerned with group-based analyses, although individual differences were observed in children's production [1]. This study investigates how differences in children's production are received by adults and what causes the differences. Factors including children's comprehension, verbal intelligence and musicality are considered.

[2] argued for a production-comprehension asymmetry in children's prosodic focus-marking, suggesting that children better in comprehension may also be better in production. Verbal intelligence (the ability to analyse information and solve problems using language-based reasoning) appears to be positively correlated with language development in general [3] and prosodic focus-marking in Chinese-speaking 5-year-olds in particular [4]. It remains to be seen whether this is applicable to Dutch-speaking children. Further, better musical abilities predict better pitch-related prosodic abilities in language production and perception [5]. But there is also evidence for lack of connection between musical and prosodic abilities [6]. It is thus not clear whether musicality is related to the ability to use prosody in focus-marking.

Seventy-five typically-developing Dutch-speaking 4- and 5-year-olds (mean age: 5;3) participated in this research. In the production experiment, SVO sentences with focus on subject-NPs (initial-focus), and object-NPs (final-focus) were elicited as responses to who/what-questions in an interactional and controlled sitting. Usable full-sentence responses and corresponding questions were combined into context-response dialogues and were subsequently evaluated for contextual appropriateness of the prosody on a 5-point scale by three trained judges. Second, the children participated in a comprehension experiment, in which they judged animals' answers to who/what-questions raised by a boy about some pictures [2]. Their reaction times (RT) to sentences with appropriate prosodic focus-marking and sentences with inappropriate prosodic focus-marking were recorded. Verbal intelligence was assessed via the PPVT-III-NL [7]. Musical abilities were assessed via the PMMA test [8].

Verbal intelligence and musicality scores were obtained following standard procedures. A 'production' score was computed for each child in each focus condition by averaging the scores of available responses. The production scores ranged from 1 to 5 (std: 0.87), indicating clear perceptual consequence of non-adultlike prosody and substantial individual variation. A 'comprehension' score was computed for each child by calculating the ratio between the mean log-transformed RT in the inappropriate-prosody condition and that in the appropriate-prosody condition in each focus condition. Linear regression analysis was conducted on the data from the children who had all four kinds of scores. The models revealed that only musicality was a significant predictor, accounting for 22% of the variation in production in the final focus condition (r=.47, p = .03).

Our results thus provide evidence for the connection between musical and prosodic abilities. Further, they suggest that verbal intelligence may have language-specific effects on acquisition of prosodic focus-marking. Its relevance to Chinese children's prosodic focus-marking points to a possible connection between a larger vocabulary and better control of post-lexical use of pitch. Finally, there may be some degree of independence in producing prosody and comprehending prosody.

References

[1] Chen (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *J. of Child Language*.

[2] Chen (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*.

[3] Groth-Marnat, G. (2009) *Handbook of psychological assessment* (5th ed.). New York: Wiley.

[4] Chen (2012). Individual differences in children's focus marking: verbal intelligence and theory of mind. Paper presented at the 5th International Conference on Tone and Intonation in Europe. Oxford, 2012.

[5] Patel, A.D. & Iversen, J.R. (2007) The linguistic benefits of musical abilities. *Trends in Cognitive Sciences,* 11, 369-372.

[6] Peretz, I. (2009) Music, language and modularity framed in action. *Psychological Belgica*, 49(2-3): 157-175.

[7] Dunn, L. M. and Dunn, L. M. (2005). *Peabody Picture Vocabulary Test-III-NL, Nederlandse versie door Liesbeth Schlichting*. Harcourt Assessment B.V., Amsterdam.

[8] Gordon, E. (1986). Primary Measures of Music Audiation (PMMA) (K-Grade 3). GIA Publicaitons, Inc.

Effects of dyslexia and musicality on the perception of rhythmic structure

Natalie Boll-Avetisyan[1], Anjali Bhatara[2], & Barbara Höhle[1]
[1] University of Potsdam, [2] Université Paris Descartes

It has often been observed that humans tend to group tone or syllable sequences that vary in intensity as strong-weak, whereas they group sound sequences that vary in duration as weak-strong (e.g., Woodrow, 1909). Likewise, across languages and cultures, word- and (musical and linguistic) phrase-initial stress is marked by intensity, whereas final stress is marked by duration. Due to this parallel, Hayes (1995) proposed a general auditory principle that guides rhythmic perception, the Iambic/Trochaic law (Hayes, 1995).

Recent studies indicate that language experience affects rhythmic grouping. For example, French speakers have weaker grouping preferences than German speakers, especially if the stimuli are intrinsically complex (e.g., Bhatara et al, 2013). Following the stress "deafness" hypothesis (e.g., Dupoux et al., 1997), the authors suggest that French speakers, lacking word stress in their native language, encode the streams at a lower level than German speakers, who might draw on abstract prosodic representations, as their native language has lexical stress.

This account can be put under test by examining dyslexics, who are assumed to have higher-order phonological representation deficits (Snowling, 2000; Ramus, 2003). Interestingly, some aspects of phonological processing are enhanced in dyslexics if they are trained musicians (e.g., Bishop-Liebler et al., 2014), suggesting they benefit from cross-domain transfer. Hence, we hypothesized that dyslexics have weaker grouping preferences than non-dyslexics, and that musicality affects rhythmic speech perception.

We tested 14 German-speaking adult dyslexics and 14 controls (so far; goal: 20/group). We used Bhatara et al.'s (2013) grouping task: they had to listen to nonsense speech sequences in which syllables alternated in duration (e.g., *...boomuzeeli...*), intensity (e.g., *...BEluMOle...*), or neither (e.g., *...bezilemo...*), and had to indicate whether they perceptually grouped syllables as strong-weak or as weak-strong. Afterwards, they completed standardized tests on their musical receptivity for melody and rhythm (MET; Wallentin et al., 2010) and on other cognitive abilities.

A mixed logit model revealed that dyslexics gave more weak-strong responses in the duration condition, and more strong-weak responses in the intensity condition than expected by chance. In the control condition, they had no significant preference for a grouping. A second model included group, condition and musical receptivity for rhythm and melody as fixed factors. For condition, we coded a contrast comparing grouping preferences between (a) intensity- and duration-varied stimuli, and (b) between duration-varied and control stimuli. Results revealed that dyslexics had weaker grouping preferences than non-dyslexics ((a) $p < .001$, and (b) $p < .05$). There was no effect of melody perception ability, but grouping preferences were enhanced if participants were better at discriminating musical rhythm ($p < .001$).

The results meet the assumptions of the Iambic/Trochaic law as a domain-general auditory principle, as grouping preferences are influenced by musical receptivity for rhythm, even if transfer between music and language is selective. Moreover, dyslexics' grouping preferences were weaker than those by non-dyslexics. This is expected in light of their phonological deficit: it seems that dyslexics—just like the French in prior studies—do not benefit from higher-order prosodic representations when encoding rhythmic structure.

References

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T. & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *Journal of the Acoustical Society of America*, 134, 3828–3843. Bishop-Liebler, P., Welch, G., Huss, M., Thomson, J. M., & Goswami, U. (2014). Auditory Temporal Processing Skills in Musicians with Dyslexia. *Dyslexia*, 20, 261–279.

Dupoux, E., Pallier, C., Sebastian-Galles, N., and Mehler, J. (1997). A destressing 'deafness' in French. *Journal of Memory and Language*, 36, 406–421.

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

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13, 212–218.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Oxford, England: Blackwell.

Woodrow, H. (1951). Time perception. In: *Handbook of Experimental Psychology*, edited by S. Stevens. Oxford : Wiley and Sons.

Rhythmic context affects on-line ambiguity resolution in silent reading

Mara Breen & Johanna Kneifel (Mount Holyoke College)

A recurring question in reading research is whether the features of the inner voice generated during silent reading are similar to those of overt reading. Under the *implicit prosody hypothesis* (Fodor, 1998), prosodic factors that affect ambiguity resolution in overt reading can also affect disambiguation in silent reading. For example, Breen & Clifton (2011) showed that syntactic reanalysis was slowed when it required simultaneous stress reanalysis. In the current study, we investigated whether a sentence's rhythmic structure affects on-line disambiguation of words with alternating stress patterns.

Thirty-four participants' eye movements were recorded while they read exchanges below. The pre-target sentence (1) established wide focus on the target sentence. The target sentence (2) always had the form *It's that the A/N N/V... A/N* had one or two (trochaic) syllables and could be interpreted as either an adjective or a noun (*dumb, stupid*). *N/V* was a stress-alternating noun-verb homograph (*PERmit*=noun; *perMIT*=verb). According to Celex (Baayen, Piepenbrock, & van Rijn, 1993), the *A/N'*s did not differ significantly as a group according to how often they were resolved as an adjective vs. a noun (t < 1).

We manipulated two factors: (a) The disambiguation of the target, and (b) the number of syllables in the *A/N*. The disambiguating region provided a continuation of the sentence which was consistent with an ADJ-NOUN interpretation of the target region (*A/N N/V*) (as in (2a, 2b)), or with a NOUN-VERB interpretation of the target region (2c, 2d). We predicted that readers would interpret the target as ADJ-NOUN, leading to difficulty when they encountered information consistent with the NOUN-VERB reading at *the spending...* in (2c,d). Crucially, we predicted that reanalysis would be more difficult when the global rhythmic pattern of the sentence predicted stress on the first syllable of *permit*, as in (2d), because both the syntactic and rhythmic information would both be consistent with a noun interpretation.

As predicted, readers initially interpreted *dumb/stupid permit* as an ADJ-NOUN and had to revise their initial interpretation, as evidenced by a main effect of disambiguation on go-past times on the disambiguating region (Region 5), t = 5.3; however, this effect interacted with syllable number such that reading times on Region 5 were even longer when the target was preceded by a two-syllable *A/N*, t = 2.52. This interaction was evident in second-pass times on Reg 3 (t = 2.65) and Reg 4 (t = 1.8) indicating that readers regressed to resolve the ambiguity.

These results demonstrate that syntactic reanalysis is facilitated when the metrical pattern of the sentence is consistent with the metrical reanalysis of the ambiguous word. The results suggest that readers use rhythmic information to make real-time predictions about syntactic category information. Moreover, they add to a growing body of work demonstrating that metrical information is represented on-line during silent reading (e.g., Kentner, 2012).

1. What's the problem with the estate?

	GP (Reg5)	SP (Reg3)
2a. 1-SYLL/NOUN:		
It's **that** the$_2$| **dumb**$_3$| per**mit**$_4$| is a nuisance for$_5$| everyone$_6$.	1069 (63)	65 (10)
2b. 2-SYLL/NOUN:		
It's **that** the$_2$| **stu**pid$_3$| **per**mit$_4$| is a nuisance for$_5$| everyone$_6$.	909 (49)	104 (17)
2c. 1-SYLL/VERB:		
It's **that** the$_2$| **dumb**$_3$| per**mit**$_4$| the spending of their$_5$| savings$_6$.	1236 (69)	116 (14)
2d. 2-SYLL/VERB:		
It's **that** the$_2$| **stu**pid$_3$| **per**mit$_4$| the spending of their$_5$| savings$_6$.	1410 (78)	254 (23)

Phonological overlap affects word duration in discourse contexts

Loretta K. Yiu & Duane G. Watson (University of Illinois at Urbana-Champaign)

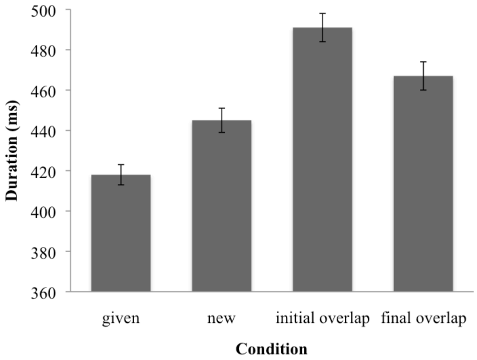
Speakers tend to lengthen new or unpredictable words and shorten words that are repeated or predictable (e.g., Aylett & Turk, 2004; Fowler & Housum, 1987; many others). One explanation for these duration effects is that reduction is related to ease of lexical access (e.g., Bell et al., 2009; Lam & Watson, 2010). Words that are retrieved more quickly are produced with shorter durations. If lengthening is linked to lexical access, however, why would a speaker benefit from lengthening a word once articulation has begun and lexical information has presumably already been accessed?

One possibility is that lengthening words benefits phonological encoding processes at points of complexity within a word. If phonological selection is a serial process (Sevald & Dell, 1994; O'Seaghdha & Marin, 2000), and activation spreads interactively between lexical and phonological representations, then words that overlap initially (e.g., PICK-PIN) should be more difficult to produce than words that overlap finally (e.g., PICK-TICK). For example, in the case of initial overlap, a speaker who intends to say PICK will initially produce a P, which will send feedback to both PICK and PIN. These two lexical representations activate their respective phonological representations, resulting in interference that will last throughout the word, slowing articulation. In contrast, for final overlap, interference occurs relatively late, resulting in less difficulty. Thus, initial overlap should lead to greater lengthening than final overlap. These overlap effects have been modeled computationally (Sevald & Dell, 1994; Watson et al., in press) and observed in word repetition tasks in which speakers repeat two words rapidly. However, it is unclear whether this production-based account of lengthening can account for lengthening in discourse contexts.

The present study examined whether speakers' durational choices in a discourse context reflect the production processes involved in phonological encoding. Fifty-two native English speakers were shown a 2x2 display of four images and were asked to describe a shrinking and flashing event occurring in succession in each trial. We manipulated the location of phonological overlap with a previously articulated word (prime) to determine whether the type of overlap affects the duration of a target word in a sentence. The critical item was the noun in the second utterance. Examples from the four conditions are given below:

a) *Initial overlap*: The beetle shrinks. The beaker flashes.
b) *Final overlap*: The speaker shrinks. The beaker flashes.
c) *Given*: The beaker shrinks. The beaker flashes.
d) *New*: The apple shrinks. The beaker flashes.

Speakers lengthened target words to a greater extent when the words overlapped initially with their primes than when they overlapped finally (*t*=-3.08, *p*=.002). Partial overlap also led to longer durations than when the target word did not share any overlap with its prime (*t*=5.81, *p*<.001), and repeated words led to the shortest durations overall (*t*=6.79, *p*<.001). The duration difference between the overlap conditions is in line with predictions of serial phonological competition models. The fact that lengthening corresponds with phonological overlap suggests that variation in duration is linked to difficulty in phonological planning.



Cross-phrase tonal patterns cue boundary strength (variably)

Alejna Brugos & Jonathan Barnes (Boston University)

Systematic variation in pitch and timing cue prosodic boundary strength, but the relationship among cues is complicated. Research has yielded variable and sometimes contradictory results as to relative weighting of cues in boundary strength perception, supporting hypotheses of pitch-timing cue-trading relationships [1, 3, 6, 2]. Duration alone can cue grouping when pitch is neutral [8, 9], yet pitch can cue grouping when duration is neutral [5].

Many pitch cues examined in boundary studies can be considered in terms of pitch continuity; bigger discontinuities (i.e., reset) mark stronger boundaries (eg. [7,4]). However, in House's (1990 [5]) series of experiments exploring pitch cues in grouping, results for several patterns cannot be straightforwardly attributed to perceived discontinuity: groupings appeared to be individuated by tonal shapes forming coherent contours across the component phrase-sized units. Figure 1 shows tonal configurations which strongly cued grouping in sequences of five digits. These patterns resemble in shape two patterns frequently attested over a single phrase: the rise-fall, and the fall-rise.

A new experiment seeks to explore whether comparable cross-phrase tonal patterns likewise cue grouping in American English, and to examine their cue strength in interaction with timing. Stimuli (Figure 2) were sequences of 5 repetitions of *nine*, resynthesized from a 460-*ms* base file. Files were concatenated with 5 pitch patterns: two predicted to cue 3-2 grouping ("rise-fall-3-2" and "fall-rise-3-2") and two 2-3 grouping ("rise-fall-2-3" and "fall-rise-2-3"). The 5th pitch pattern ("neutral"), a steady fall across the digits, was included as a baseline. Pauses between each of the 5 digits were 200 *ms* in the neutral timing condition, and lengthened either following the 2nd or 3rd digit. The 5 pitch patterns were presented with each of 13 timing patterns, for 4 repetitions, for a total of 260 randomized trials per subject. Subjects (N=34) were asked to indicate whether each presented string was "grouped" as 999-99 ("3-2-grouping") or 99-999 ("2-3-grouping").

Results from 8840 trials (Figure 3A) showed that rise-fall patterns strongly cued predicted groupings: The predicted "rise-fall-3-2" pattern indeed cued more "3-2-grouping" responses (65% over all time steps) than the predicted "rise-fall-2-3" pattern (37%). Suggestive of cue trading, this effect was strongest with ambiguous timing cues (i.e., where pauses were equal) yielding 76% responses "3-2-grouping" for the predicted "rise-fall-3-2" pattern, to only 31% for "rise-fall-2-3". Further, this difference persisted when timing patterns strongly cued the opposite grouping (eg. Where pause durations differed by 50% to 100%).

Results for the fall-rise patterns were less straightforward. For the group as a whole, neither fall-rise pattern cued grouping differently from neutral pitch (54% for "fall-rise-2-3", 53% for "fall-rise-3-2", vs 54% for neutral). Results by individual subjects, however, revealed that some did interpret the fall-rise patterns as in House (1990) (Figure 3B), while others responded with an unpredicted opposite pattern potentially reflecting local pitch cues (Figure 3C). Results as a whole suggest that not only do pitch and timing cues differ in weighting depending on context, but that individual listeners have differing strategies in weighting these cues.

Figure 1: Pitch patterns from House (1990). The top two patterns cued a 3-2 grouping, the bottom two a 2-3 grouping. Patterns at left have in common that a generally domed rise-fall contour encompasses each of the perceived groups. Those at right have the flipped picture, with scooped/fall-rise shapes delineating the groups.

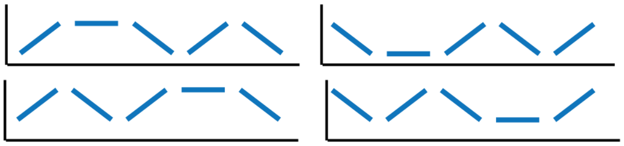


Figure 2B: A sample stimulus, from the fall-rise-2-3 pitch pattern, and with a longer pause after the 2nd pause (difference between pause 3 and pause 2 is -150 ms.)

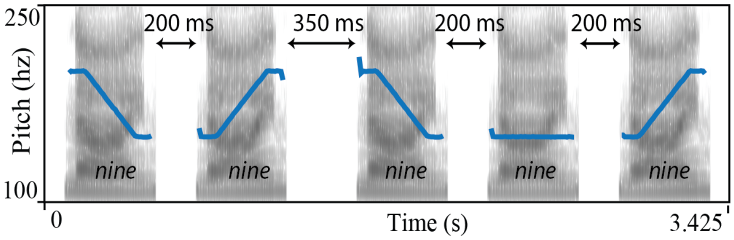


Figure 3A. Results for all 34 subjects. Rise-fall-3-2 (orange) and rise-fall-2-3 (blue) clearly separate, but fall-rise patterns (purple and green) do not differ from neutral (black)

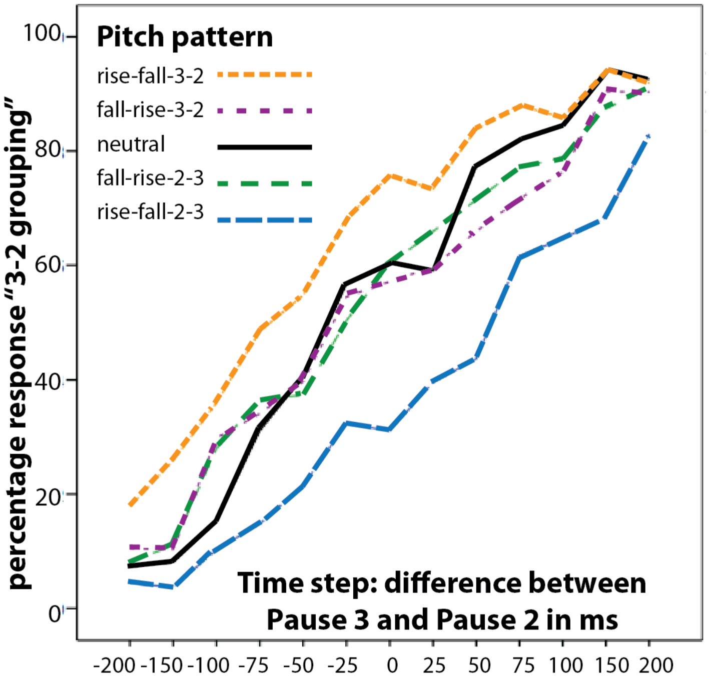


Figure 2A. Stimuli: Pitch patterns (crossed with continuum of pause duration manipulations to pause 2 & pause 3)

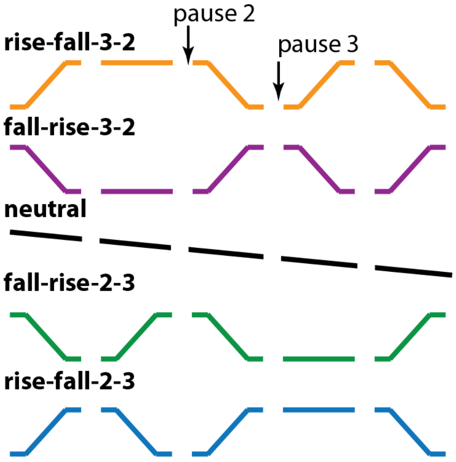


Figure 3B: Results for 10 subjects showing predicted responses for fall-rise patterns (purple and green lines)

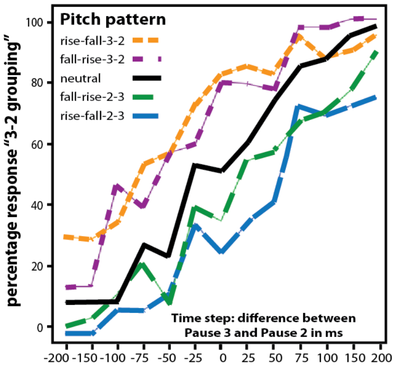
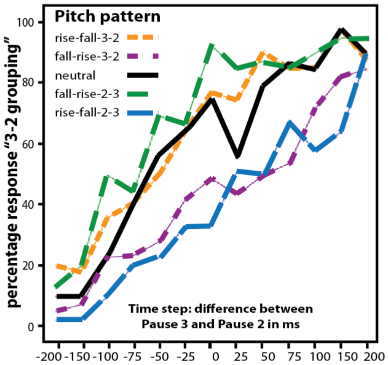


Figure 3C: 10 subjects showing opposite pattern from predicted responses for fall-rise patterns (purple and green lines generally reversed from above)



[1] Beach, C. (1991). "The interpretation of prosodic patterns..." *J. of Memory & Lang.*, **30**(6), 644–663. [2] Brugos, A. & Barnes, J. (2014) "Effects of dynamic pitch...," *Sp. Prosody 7*, pp. 388-392. [3] Cumming, R. (2011). "The interdependence of tonal and durational cues..." *Phonetica* **67**, 219–242. [4] Féry, C. & Truckenbrodt, H. (2005). "Sisterhood and tonal scaling." *Studia Linguistica*, 59(3). [5] House, D. (1990). *Tonal Perception in Speech*. Lund Univ. Press. [6] Jeon, H. & Nolan, F. (2013). "The role of pitch and timing cues in the perception of phrasal grouping…" *JASA*, **133**(5), 3039-3049. [7] Ladd, D. (1988). "Declination 'reset '...." *JASA*, 84, 530. [8] Scott, D. (1982). "Duration as a cue…" *JASA*, **71**(4), 996–1007. [9] Wagner, M. & Crivellaro, S. (2010). "Relative Prosodic Boundary Strength..." *Sp. Prosody 5*.

SET it up: Speaker knowledge and intonational contours in a natural production task

Sarah A. Bibyk, Christine Gunlogson, & Michael K. Tanenhaus (University of Rochester)

Researchers have long hypothesized relationships between intonational categories and speaker intentions such as questioning, while acknowledging that such relationships vary greatly by discourse context, speaker, and linguistic content. For example, so-called *declarative questions* are commonly associated with rising intonation, but the actual relationship is more complex[1]. This variability poses challenges for the study of intonational meaning, particularly within natural corpora. Without prior understanding of how various discourse factors interact with intonation, we risk circularity when trying to use those factors to identify speaker intentions. In particular, annotation schemes for discourse[2,3] that take intonation into account when classifying questions cannot then be turned around and used to test the association between intonation and questioning.

To provide an alternative, we developed a "targeted language game"[4] allowing for the production of naturally variable speech, together with an intonation-independent means of labeling speakers' intentions as questioning or non-questioning. Using this paradigm, we show that tracking which interlocutor is a better information source predicts intonation patterns in declarative utterances.

Pairs of participants played a cooperative card game based on the commercial game SET[5] while separated from each other by an occluder. For some cards, each player had visual access to features the other lacked, necessitating information exchange. Utterances that mentioned aspects of cards which speakers could not see (**partner** visual access) were classified as information-seeking, and therefore questions. Conversely, utterances that mentioned aspects of cards speakers *could* see (**own** visual access) were classified as providing rather than requesting information. Following Gunlogson[1], we predicted speakers would produce declarative questions in situations where they had *some* evidence for the information they were putting forward, but the addressee was ultimately a better source (e.g. in the case of a previously discussed card aspect privileged to the addressee).

We recorded productions from 11 pairs for a total of ~7 hours of data. In line with our predictions and replicating previous research[6], speakers overwhelmingly used wh-interrogatives when talking about their **partner**'s cards as opposed to their **own** (Figure 1). Although a large portion of wh-interrogatives were produced with falls, a substantial number had rising intonation, indicating our game generated productions that vary intonationally. We next determined that speakers do also use declarative syntax when talking about their partner's cards, even more frequently than polar-interrogatives (Figure 2). Thus, our game successfully elicited declarative utterances even when speakers discussed items to which they did not have visual access. Finally, we found that in declarative utterances speakers were more likely to use rises as opposed to falls when talking about their partner's card, and more likely to use falls as opposed to rises when talking about their own (Figure 3).

In summary, we have successfully constructed a paradigm to examine the distribution of intonational contours in declarative utterances and how these depend on speaker knowledge. Extending this methodology, we plan to investigate how other cues to questioning interact with intonation, as well as how discourse factors may affect the speaker's choices, in order to understand how and why intonation varies with the context.

Figure 1: Wh−interrogatives by source

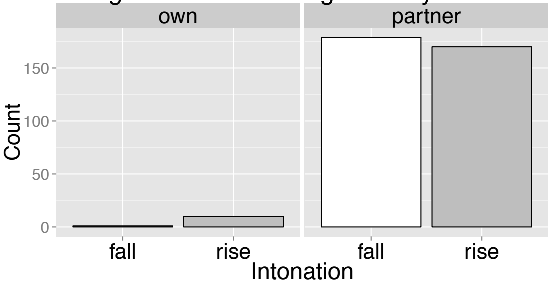


Figure 2: Utterances about partner's cards

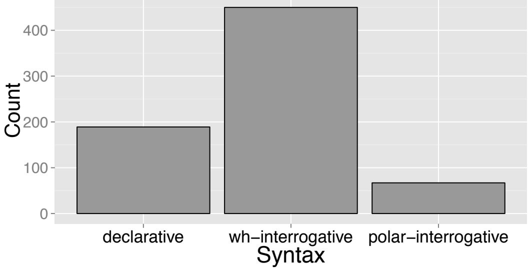
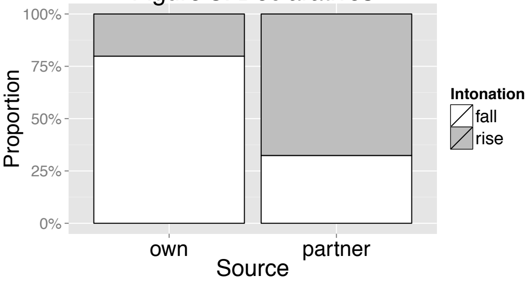


Figure 3: Declaratives



References

[1] Gunlogson, C. (2008). A question of commitment. Belgian Journal of Linguistics, 22(1), 101-136.
[2] Core, M., & Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In AAAI fall symposium on communicative action in humans and machines, 28–35. Citeseer.
[3] Jurafsky, D., Shriberg, E., & Biasca, D. (1997). Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. Institute of Cognitive Science Technical Report, 97-102.
[4] Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: a targeted language game approach. Cognitive Science, 32(4), 643–684.
[5] Set Enterprises Inc., (2014). Retrieved from www.setgames.com.
[6] Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. Cognition, 107(3), 1122–1134.

Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses

Kiwako Ito, Rory Turnbull, & Shari Speer (Ohio State University)

In the past decade, accumulating experimental evidence has confirmed that listeners interpret distinctive prosodic prominence as a cue to consider an alternative discourse entity as soon as the acoustic cue becomes available [1, 2, 3]. The generalizability of those findings must be evaluated with caution, however, because (1) experiments generally make use of highly controlled laboratory speech with exaggerated pitch excursions, and (2) the range of pitch contour variation for expressing contrast in natural speech is largely unknown. This study investigates whether pitch contours from spontaneous speech that differ distinctly from laboratory speech in their acoustic properties evoke responses comparable to those reported in the abovementioned experimental work.

The experiment compared laboratory speech with clear F0 excursions to a voice from a naïve female participant in a past study in which she gave spontaneous instructions to decorate Christmas trees [4]. Both speech types were blindly ToBI-annotated. A prominence rating experiment has shown that, despite the lack of distinctive local F0 excursions in this voice, naïve listeners assigned prominence to the adjectives when the following nouns had relatively lower F0 [5]. Naïve listeners assigned prominence for adjectives in [adjective noun] pairs more frequently for [L+H* no-acc] than for [H* !H*]. The present study used a tree decoration task similar to [2] to measure listeners' eye movements while they heard contrastive (blue drum → BROWN/brown drum) and non-contrastive (green candy → CLEAR/clear house) sequences with either [L+H* no-acc] or [H* !H*] pitch contour.

Data from 40 science museum visitors (Age: 19-65) showed a clear effect of the pitch contour type in the spontaneous speech even without the extreme F0 rise and fall typical of laboratory-generated L+H*. Just like the laboratory speech, [L+H* no-acc] contour led to faster detection of the target ornament (e.g., brown drum) as compared to [H* !H*] contour in the contrastive sequence. In addition, [L+H* no-acc] led to the initial incorrect looks to the previously mentioned ornament (i.e., contrastive competitor: e.g., candy) in the non-contrastive sequence.

The two speech types, however, yielded a difference in the timing of fixations: While extreme F0 excursion within the adjective for L+H* in the laboratory speech led to earlier detection of contrastive target and earlier incorrect looks to the contrastive competitor (initiated in the mid point of the noun), the lack of F0 excursion in the adjective in the spontaneous speech led to later fixation increase (initiated toward the end of the noun). With these findings, we confirm that pitch accents or intonation contours are abstract interpretational units that may be realized with a range of allophonic tunes. Listeners seemed to tune to the speakers' idiosyncratic use of prosodic cues to identify the pragmatic intent of the utterances, although the degree of complexity and clarity of tonal cues may affect the timing of responses. Annotation systems must capture abstract phonological categories that are consistent with pragmatic distinctions perceived by listeners across a range of phonetic implementations.

References

[1] Weber, A., Braun, B., Crocker, M. W. (2006). Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Language and Speech*, *49* (3), 367-392.
[2] Ito, K. & Speer, S. R. (2008). Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*, 541- 573.
[3] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133*, 335-342.
[4] Ito, K & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In P. Augurzky & D. Lenertova (eds), *Methods in Empirical Prosody Research*. (pp.229-257) Mouton de Gruyter.
[5] Turnbull, R., Royer, A., Ito, K. & Speer, S. R. (2014). Prominence perception in and out of context. *Speech Prosody* 7. Dublin, May.

Prosodic disambiguation in Turkish: Evidence from phoneme restoration

Nazik Dinctopal-Deniz (Bogazici University) & Janet Dean Fodor (The Graduate Center, The City University of New York)

Background & Purpose: Kjelgaard and Speer [1] found that inappropriate prosodic contours caused more processing difficulty in English early closure (EC) structures than late closure (LC) structures. We consider two explanations for this asymmetry: (i) When prosodic phrasing is uninformative/misleading about syntax, the parser resorts to the classic *syntactic* LC strategy; (ii) Prosodic breaks flanking short constituents are treated as more informative about syntax than breaks flanking longer constituents, as per the Rational Speaker Hypothesis (RSH) [2].

To investigate these hypotheses, we tested an LC/EC Turkish ambiguity in two listening experiments employing the phoneme restoration (PR) paradigm which has been shown to be sensitive to listeners' use of prosodic cues in ambiguity resolution in Bulgarian [3]. We tested whether PR shows prosody-sensitivity in Turkish, and whether it would reveal RSH effects.

Materials & Procedure: The Turkish ambiguity (see below) had a morphological ambiguity between subject or object interpretation of the second DP. A noise burst replaced the disambiguating morpheme that marked the verb as transitive or intransitive, creating a global ambiguity. Modifiers were inserted to vary the phrase lengths. Expt 1 items had lengthened subject but short VP. Expt 2 items had short subject and lengthened VP. All sentences were pronounced with LC, EC or neutral prosody. Following a sentence, listeners saw a visual probe word (the verb, complete with all morphemes, unambiguously transitive or intransitive), which was congruent, incongruent, or compatible with the prosody of the sentence they heard. They were asked 'Did you hear this word?'.

Results: Participants responded 'yes' more often to the congruent probe than the incongruent probe, with the compatible probe in between (*p*'s < .001). Evidently, they had restored the missing phonemes (creating either LC or EC syntax) by reference to the prosodic phrasing. RT data showed a similar pattern (*p*'s < .001).

In accord with hypothesis (i), there was an LC advantage in conditions where prosody was not helpful. With neutral prosody (compatible probe), in both experiments, the observed LC advantage would result from a syntactic LC strategy emerging when not masked by overt prosodic cues. LC advantage in the incongruent probe condition in Expt 2 can also be attributed to syntactic LC bias, functioning as a default/reanalysis strategy when prosody is present but misleading. The incongruent probe condition in Expt 1 showed an EC advantage. Though unanticipated, this can be ascribed *post hoc* to listeners being unable to make use of the prosodic boundary in the LC items because the noise-replaced word occurred immediately (too soon) after it (see also [3]).

Regarding hypothesis (ii), neither 'yes' responses nor RTs showed any response differences related to the phrase-length contrasts across the two experiments. This differs from a prior study with similar materials in a 'got it' task, and needs explaining. We suggest that estimating prosodic informativeness by consulting phrase lengths, as per RSH, is most valuable to the parser when sentences contain both prosodic and morpho-syntactic cues to syntactic structure, which could potentially conflict.

Spoken Sentence*	**Visual Target**
a. LC prosody	LC / EC
(Yaklaşık) yedi öğrencinin psikoloğu || (oldukça) sev i sandım.	sevildi / sevdiğini
b. EC prosody	LC / EC
(Yaklaşık) yedi öğrencinin || psikoloğu (oldukça) sev i sandım.	sevildi / sevdiğini
c. Neutral prosody	LC / EC
(Yaklaşık) yedi öğrencinin psikoloğu (oldukça) sev i sandım.	sevildi / sevdiğini
(Nearly) seven student$_{\text{-GEN}}$ psychologist$_{\text{-3SG.POSS/ACC}}$ (much) like- think$_{\text{-1SG}}$	like$_{\text{-PASS}}$ / like$_{\text{-FN-ACC}}$

LC: 'I thought that the psychologist of (nearly) seven students was (much) liked.'
EC: 'I thought that (nearly) seven students liked the psychologist (much).'

* || indicates prosodic boundary

[1] Kjelgaard, M. M. & Speer, S. R. (1999). *Journal of Memory and Language, 40*(2), 153-194.
[2] Clifton, C., Carlson, K., & Frazier, L. (2006). *Psychonomic Bulletin & Review, 13*(5), 854-861.
[3] Stoyneshka, I., Fodor, J. D., & Fernández, E. M. (2010). *Language and Cognitive Processes, 25*(7), 1265-1293.

****

Prosodic juncture strength and syntactic constituency in Connemara Irish

Emily Elfner (University of British Columbia)

Introduction: This paper discusses the results of a production experiment on Connemara Irish (CI), designed to investigate the correlation between syntactic constituent structure and the relative strength of the junctures found at prosodic boundaries in transitive (VSO) sentences, which syntactically have the structure [V [[DP$_{\text{SUBJ}}$] [DP$_{\text{OBJ}}$]]] (e.g. McCloskey 2011). It was found that, in apparent contradiction to the hypothesis that syntactic constituency and prosodic boundary strength should correlate (Lehiste 1973; Wagner 2005), the strongest prosodic juncture in VSO sentences occurs between S and O rather than V and S. However, rather than assuming that the unexpected location of the prosodic juncture is representative of a mismatch between syntactic and prosodic structure, I propose that patterns of juncture strength in CI are compatible with the hypothesis that syntactic and prosodic structure are isomorphic, under the assumption that juncture strength is influenced by both syntactic and processing-based factors.

Experiment: The production experiment was designed to investigate the prosodic properties of transitive (VSO) sentences in CI. The experiment had a 3x3 design which systematically varied the relative **length** of the subject and object between 1 and 3 words, along the following schema:

(1) Experiment design (brackets indicate constituency but not complete syntactic structure)

Argument length	O=1	O=2	O=3
S=1	V [N] [N]	V [N] [NA]	V [N] [NAA]
S=2	V [NA] [N]	V [NA] [NA]	V [NA] [NAA]
S=3	V [NAA] [N]	V [NAA] [NA]	V [NAA] [NAA]

DP arguments consisted either of Noun-Adjective constructions (as above) or possessive constructions, which were systematically crossed across items. Eight native speakers of CI read 108 sentences each (12 items x 9 conditions).

Results: Juncture strength was measured at two locations: between V and the first word of the subject (**J1**) and between the last word of the subject and the first word of the object (**J2**): V $_{J1}$ S $_{J2}$ O. Three acoustic measures are discussed: word duration of V for J1 and the last word of the subject for J2, the likelihood of a pause at each juncture position, and the combined duration of the word and the following pause (if present).

(2) Results of the three acoustic measures

Acoustic Measures	**J1**	**J2**

Word duration	S=1 > S=2, S=3
	No difference with length of O
		S=3 > S=2 > S=1
		O=3 > O=2 > O=1

Likelihood of pauses	S=3 (14.7%) > S=2 (7.1%) > S=1 (3.9%)
	No difference with length of O
		S=3 (52.4%) > S=2 (39.7%) > S=1 (32.3%)
		O=3 (51.3%) > O=2 (41.5%) > O=1 (31.4%)

Duration of target word + pause duration	S=1 > S=2, S=3
	No difference with length of O
		S=3 > S=2 > S=1
		O=3 > O=2 > O=1

Discussion: These results show that (a) there is more likely to be a juncture at J2 than at J1, and that (b) the relative strength of J2 (but not J1) increases with the length of both subject and object. Based on the syntactic bracketed structure alone, this finding is surprising: assuming that S and O form a constituent in CI, we would expect to find a relatively stronger juncture following V than following S. My proposal is that the **presence** of the juncture at J2 is motivated by the phrase-final status of the preceding word: the subject is contained within a syntactic phrase (DP), while the verb is non-phrasal as the head of the larger clausal constituent, and thus not phrase-final. However, because the relative strength of J2 is conditioned by the length/complexity of both upcoming and recently uttered material, we can hypothesize that juncture **strength** is thus influenced by processing load and production planning (as a function of length and/or complexity), rather than a straightforward interpretation of syntactic/prosodic bracketing (Ferreira 1991, 1993; Watson & Gibson 2004).

References

Ferreira, Fernanda. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language* 30:210-233.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100:233-253.

Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa* 7:107-123.

McCloskey, James. 2011. The shape of Irish Clauses. In A. Carnie (ed.) *Formal Approaches to Celtic Linguistics*. Cambridge: Cambridge Scholars Publishing.

Wagner, Michael. 2005. Prosody and recursion. Doctoral dissertation, MIT.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19:713-755.

Semantic expectation affects prosodic prominence perception

David Lutz & Sam Tilsen (Cornell University) del82@cornell.edu

Though prosodic prominence is reasonably well-understood acoustically as a combination of increased $F_0$and F_0range, duration, and intensity, its perception is also affected by top-down, expectation-based factors like token frequency and local repetition count (Cole et al. 2010). Since contrastive focus is a well-studied semantic phenomenon that has demonstrated effects on prosodic prominence (Katz & Selkirk 2011), the expectation of contrastive focus may affect prominence perception.

Previous work supports two competing hypotheses. Results reported in Wagner 2005 *et seq*. suggest that listeners' perceptions tend toward their expectations, such that listeners will report greater prominence when they expect the target to be more prominent than when they don't. In contrast, Parker et al. (2012) found that when asked to rate the loudness of sounds, participants' responses tended away from their expectations, i.e. participants who expected loud sounds rated stimulus sounds as softer than participants who had no such expectation rated the same stimuli. The resulting gain control model of audition, by which listeners adjust the "gain" of their auditory system based on their expectations, predicts that participants report lower prominence when they expect a word to be focused than when they don't.

In a perceptual experiment, using written contexts and questions preceding audio stimuli (c.f. Breen et al. 2010), we caused participants to expect either standard declarative prosody (normal context) or contrastive focus on a target word (focus context) in stimulus utterances. We then manipulated whether the synthesized stimulus utterances had normal declarative prosody (normal stimulus) or contained focus on the target word (focus stimulus), in which the focused element was spoken with increased pitch and duration, and subsequent elements were deaccented. Participants therefore responded to four different experimental conditions in a two-by-two factorial design.

16 participants responded to 62 target stimuli each, balanced across conditions and randomized across participants, and 90 distractors. Target stimuli all had the same syntactic and prosodic structure, while distractors varied prosodic structure and target word location. Participants rated the prominence of a target word in an utterance on a visual analog scale, marked with "Not at all prominent" on the left and "Extremely prominent" on the right, and containing no other markings. Responses were recorded as integers between 0 and 100.

A two-factor ANOVA confirmed that stimulus (normal vs. focus stimulus) is a highly significant predictor of prominence rating ($F = 437.04$; $p \ll 0.001$) as expected. Context (normal vs. focus context) is also highly significant in prominence perception ($F = 33.37$; $p \ll 0.001$). No significant interaction effect was found. Participants used the visual analog scale effectively, though all participants used the higher half of the scale much more than the lower half. Large differences between individual participants in the relative strength of context/expectation and stimulus/acoustic information on prominence ratings warrant further attention.

This result confirms that semantic expectation affects prosodic prominence perception, and that prominence tends toward the prominence the listener expected to hear.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7-9). 1044–1098. doi:10.1080/01690965.2010.504378.

Cole, Jennifer, Yoonsook Mo & Mark Hasegawa-Johnson. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2). 425–452. doi:10.1515/labphon.2010.022.

Katz, Jonah & Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771–816. doi:10.1353/lan.2011.0076.

Parker, Scott, Julianne M Moore, Sara Bahraini, Kathleen Gunthert & Debra a Zellner. 2012. Effects of expectations on loudness and loudness difference. *Attention, Perception & Psychophysics* 74(6). 1334–42. doi:10.3758/s13414-012-0326-8.

Wagner, Petra. 2005. Great Expectations – Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates. In *Interspeech* 2005, 2381–2384.

Top-down processing of intonational boundaries

Andrés Buxó-Lugo & Duane Watson (University of Illinois at Urbana-Champaign)

Intonational boundaries are discontinuities in the speech stream that are typically signaled by pauses, changes in F0 contour, and pre-boundary lengthening (e.g. Klatt, 1975; Pierrehumbert and Hirschberg, 1990; Turk & Shattuck-Hufnagel, 2007; Ladd, 2008). It is generally assumed that listeners represent prosodic boundaries and phrasing using an abstract prosodic representation (Ferreira, 1993, Turk & Shattuck-Hufnagel, 1996), but little is known about how this representation is parsed. In contrast, we know a great deal about how syntactic representations are parsed, and that listeners use prosodic information to make inferences about syntactic representations (see Wagner & Watson, 2010 for a review). However, if prosody needs to be processed in its own right, we might expect an interactive exchange between syntactic constraints and prosodic constraints in building linguistic structure: just as prosody guides syntactic parsing, syntactic knowledge might guide prosodic parsing. If this is the case, listeners may be more likely to experience hearing a boundary when it occurs at a location that is likely given the syntax. Consistent with this prediction, Cole, Mo, & Baek (2010) found that syntax is a reliable predictor of where listeners report hearing prosodic boundaries, independent of acoustics.

We explored these hypotheses using a boundary detection task. In the example below, (a) is a syntactically *inappropriate* location for a boundary and (b) is a syntactically *appropriate* location. The question was whether listeners would be more likely to report hearing a boundary at location (b) than location (a), independent of acoustic factors.

We balanced the presence of the words "green" and "frog" at each of these locations to ensure that lexical differences did not drive any of the perceived effects. We also manipulated the acoustic properties of the potential pre-boundary word by resynthesizing its duration, following pause duration, and F0 contour using PSOLA. A 9 step continuum was created. On one end of the continuum, acoustic cues were consistent with a boundary. On the other end of the continuum, acoustic cues were consistent with the absence of a boundary.

In Experiment 1, 18 participants from Amazon Mechanical Turk were presented sentences like those in (1) and (2). They judged whether a boundary was present after each word in the sentence. There were 4 different items and participants heard all 9 steps for each sentence. There was a main effect of acoustics, such that more boundaries were reported when the acoustics were consistent with a boundary (z=4.49, p<.001). Critically, there was also a main effect of syntactic type (z=-4.83, p<.001), such that across the continuum (see Figure) listeners reported a boundary more often in the syntactically appropriate location than in the inappropriate location, independent of acoustics.

The instructions in Experiment 1 included 2 example sentences that contained boundaries in syntactically appropriate locations. To rule out the possibility that the effects in Experiment 1 are the result of biases created by these instructions, a second experiment was run. Experiment 2 replicated the original results with a set of instructions that included sentences with boundaries in syntactically appropriate and inappropriate locations.

These data suggest that syntactic structure is used by listeners in building prosodic representations, and that listener expectations, along with acoustic cues, are central in determining whether a listener perceives a boundary.

1. Syntactically Appropriate After Frog: Put the **green** (a) **frog** (b) in the box.
2. Syntactically Appropriate After Green: Put the **frog** (a) that's **green** (b) in the box.

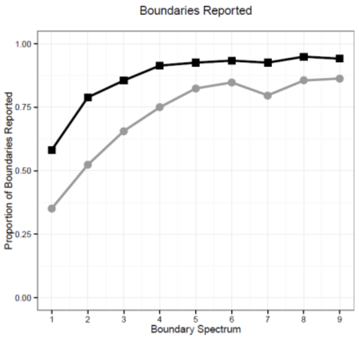


Figure: Syntactically appropriate location in black; syntactically inappropriate location in gray.

Syntactic constraints on the variability of prosodic phrasing and parenthetical placement

Aron Hirsch (MIT) & Michael Wagner (McGill University)

Syntax/prosody mismatch. The placement of prosodic boundaries seems to not always correlate with syntactic structure (see (1) for examples from the literature). We argue that constraints on the variability of prosodic phrasing provide evidence that this variation in prosodic phrasing in fact reveals variation in syntactic constituent structure, following similar ideas in Steedman (1991, et seq). However, we argue that the mechanism that derives alternative bracketings is optional rightward movement**.** An account in terms of rightward movement can predict the possibility of the phrasings in (1) (optional high attachment), and the impossibility of those in (2). In this paper, we present two experimental studies that further support this account, and distinguishes it from alternative views.

Parentheticals. Certain adverbs (*unfortunately*, *please*) can appear in different positions in the linear string, (3). To account for this, one could assume that these adverbs attach at different syntactic heights. Alternatively, these adverbs might always attach high (e.g. to TP), and constituents can appear to the right of them only by undergoing rightward movement (cf. Potts 2005, Stowell 2002). This reduces (3c) to a special case of (3b), with rightward movement of the constituent following *unfortunately*. (3c) has a signature prosody: *unfortunately* is followed by a strong boundary (in addition, unfortunately is always deaccented in these positions given rise to the intuition of a preceding boundary). Whereas complex VPs in English are usually prosodically flat (Taglicht 1984, Wagner 2005), phrases which attach high outside the VP (e.g. extraposed relative clauses) are preceded by a strong boundary. The boundary following *unfortunately* in (3c) is as expected if the following constituents attach high following rightward movement.

Multiple parentheticals as a test case. An account in terms of rightward movement makes a crucial prediction for cases where *multiple* constituents move to the right of a high adverb (as in (4c)): since each constituent is at least as high as the adverb, each should be set off by a strong prosodic boundary. If the two PPs move separately above *please*, each should be preceded by a boundary (as per prediction); if they move as a constituent, there should be one boundary preceding them together. In the case of (4b) on the other hand, only one PP rightward moves, so only it should be preceded by a boundary. Participants in a production study (n=30) produced sentences like (4) in a context ensuring that the two PPs were interpreted as VP modifiers (9 item sets, data perceptually annotated for boundary placement). As predicted by the rightward movement account, in (4c) two boundaries are very likely, while this phrasing is essentially not used for (4b).

Prosodic variability in the absence of parentheticals. Even in the absence of parentheticals, there is variability in whether speakers produce prosodic cues to disambiguate certain PP-attachments (e.g. Snedeker & Trueswell 2003, Krajic & Brennan 2005). We report experimental results from a second production study showing that speakers, when they produce sentences like (5), reliably disambiguate (5a) from (5c), but not from (5b). (5a) can be expressed with a single boundary before *on the hat*, making it indistinguishable from the default prosody for (5b). We argue that the availability of this prosody for (5a) and the absence of other prosodic variation is a consequence of syntactic constraints on rightward movement. More generally, this provides further evidence that prosodic variability in fact reveals syntactic variability.

(1) Apparent mismatches between syntax and prosody from the prior literature
a. Everyone knows (||) *that this is not true*. (Taglicht 1998)
b. She gave her friend (||) *an interesting book*. (Taglicht 1998)
c. George and Mary (||) *give blood*. (Shattuck-Hufnagel & Turk 1996)

(2) Impossible phrasings according to the prior literature
a. *But [almost || *all of them] knew that*. (Taglicht 1998)
b. *[Danish || *beer] is better*. (Taglicht 1998)
c. *[George and || *Mary] give blood*. (Shattuck-Hufnagel & Turk 1996)

(3) a. **Unfortunately**, || researchers have refuted that chocolate is healthy.
b. Researchers have refuted that chocolate is healthy,(||) **unfortunately**.
c. Researchers have refuted, ,(||) **unfortunately**, || that chocolate is healthy.

(4) a. Tap the frog *with the flower on the hat*, **please**. (final *please*)
b. Tap the frog *with the flower*, **please**, || *on the hat*. (late *please*)
c. Tap the frog, **please**, || *with the flower* || *on the hat*. (early *please*)

(5) Tap the frog with the flower on the hat. (three-way ambiguous: 5a-c)
a. *Reading 1 (‘list’)*: “Tap the frog, using a flower, and tap it on the hat.”
b. *Reading 2 (‘left’)*: “Tap the frog which is holding a flower, and tap it on the hat.”
c. *Reading 3 (‘right’):* “Tap the frog by using the flower which is lying on the hat.”

Poster Sessions

Prosodic phrasing and individual differences in relative clause attachment

Jason Bishop[1], Adam Chong[2], & Sun-Ah Jun[2]
[1] CUNY, [2] UCLA

A large body of work has appealed to prosodic structure in explaining the resolution of attachment ambiguities such as *Someone shot the servant of the actress who was on the balcony.* (Cuetos & Mitchell, 1988). According to the influential Implicit Prosody Hypothesis (IPH; Fodor 1998), the presence of a prosodic boundary directly following NP1 (which groups NP2 prosodically with the RC) should favor a low attachment parsing of the RC (i.e., to NP2, *the actress*); a prosodic boundary after NP2, on the other hand, should encourage high attachment (i.e., to NP1, *the servant*). However, controlled experimental evidence for these correspondences in explicit (i.e., overtly spoken) prosody is currently lacking, as most previous investigations of explicit prosody have focused on prominence structure rather than alternations in phrasing (Schafer et al., 1996; Lee & Watson, 2011).

The present study aimed to address this gap. We present an experiment in which English-speaking listeners (N=107) made attachment decisions about ambiguous RCs in auditorily-presented sentences like the one above. These sentences varied in the location of a prosodic boundary (a L-L% in the ToBI framework) across three conditions: an *early* boundary (after NP1) condition, a *late* boundary (after NP2) condition, or a control condition that lacked any prosodic boundary. Prominence (i.e., accentual) structure, which is known to influence RC attachment, was held constant across phrasing conditions. Upon hearing a test sentence with one of the three prosodic structures, listeners made a decision regarding RC attachment, elicited via visual scenes. Finally, listeners also completed a measure of "autistic traits", as these traits have been shown to predict individual differences in sensitivity to prominence in RC attachment (Jun & Bishop, in press).

Results (based on mixed-effects logistic regression) demonstrated the following. First, listeners' attachment decisions were, overall, sensitive to prosody as predicted by the IPH: high attachments were more likely for sentences with late boundaries (p<.001), and low attachments more likely following early boundaries (p<.001). These patterns were also sensitive to individual differences in autistic traits; although only marginally significant (p=.057), the effect of late boundaries increased as autistic traits increased (Fig 1). Notably, the influence of autistic traits was weaker than in a previous study utilizing a prosodic priming methodology (Jun & Bishop, in press). We argue that the results support Jun & Bishop's claim that these traits predict differences in prosody-based parsing strategies: individuals with weaker autistic traits rely more on prominence for attachment decisions, while those with stronger autistic traits rely less on prominence. We suggest the fact that autistic traits played a larger role in the priming study than in the current one supports the notion that prominence's influence on attachment reflects a memory-based processing strategy (Lee & Watson, 2011). We discuss our results in the context of the IPH and the implications for prosody's role in sentence processing.

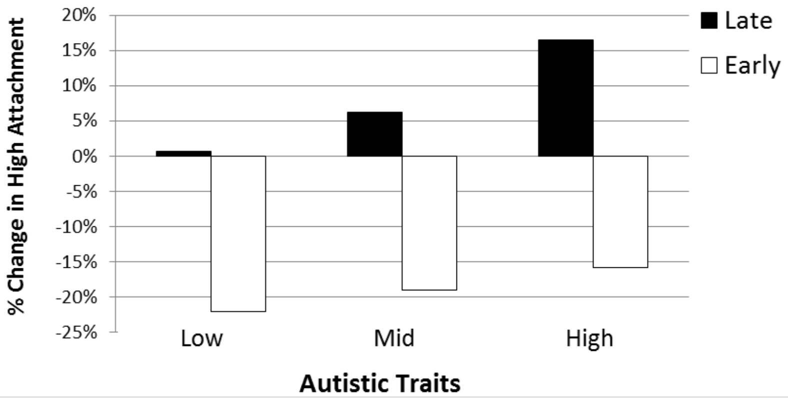


Fig 1. Change in "high attachment" responses as a function of boundary location for three different groups, based on autistic traits. The "Mid" group represents those within one standard deviation of the mean score on the "Autism Spectrum Quotient" (Baron-Cohen et al. 2001); the "Low" and "High" one sd below and above the mean.

Metrical regularity of surrounding context affects word recall

Amelia Kimball, Loretta Yiu, & Duane Watson (University of Illinois at Urbana-Champaign)

Metrical structure is formed by patterns of stressed syllables, and metrically regular instances of speech occur when stressed and unstressed syllables alternate. This is commonly found in poetry (shall **I** com**pare** thee **to** a **sum**mers **day**), but may also occur in everyday speech. Metrical regularity has been shown to have an advantage over irregular contexts in online processing and production. Listeners can track patterns of pitch movement over sentences and these patterns affect word segmentation (Dilley, 2008). Additionally, metrical structure affects online processing in silent reading (Breen and Clifton, 2010), and speakers make fewer errors in production of regular strings of non-words (Tilsen, 2011).

Evidence of a regularity advantage is surprising, however, given that reported perception of everyday speech suggests metrical regularity is not very common in conversational American English (Kimball and Cole, 2014). Furthermore, it is unclear whether this regularity advantage extends beyond online processing. Although regularity has been observed to modulate the N400 response to semantically unexpected spoken words, individuals were at ceiling (above 97% correct) in a test of semantic congruence for both irregular and regular contexts (Rothermich et al., 2012). This suggests that although there may be differences between irregular and regular sentences during processing, differences in comprehension accuracy may be minimal.

The present study examined the effects of metrical regularity on a recall memory task to test a) whether there is a psychological reality to metrical structure, and b) whether metrical structure has consequences for long-term comprehension. Recall was tested in lists of words that followed either a metrically regular or irregular pattern.

Our study used two-syllable words that had either a strong-weak pattern (stress on the first syllable, e.g. **cam**el) or a weak-strong pattern (stress on the second syllable, e.g. ca**noe**). We designed lists of eight words that either matched in stress pattern (**cup**cake, **moun**tain, **cam**el, **sand**wich...) or included a word with a conflicting pattern in the third position (**cup**cake, **moun**tain, ca**noe**, **sand**wich...). The stress pattern (strong-weak vs. weak-strong) of words in the surrounding context and in the target position were crossed in a 2x2 design. Target words were matched for frequency and imageability. Sixty online participants recruited from Amazon Mechanical Turk heard 52 lists presented in random order (13 in each condition) and were given a free recall test after every list. They were scored for number of words correctly recalled, and whether the target word was remembered.

Words are better recalled when they do not match their surrounding metrical context (z=4.77, p<.001). This oddball effect suggests that individuals are sensitive to metrical structure, and metrical structure has consequences for comprehension. Note, however, that this irregularity boost in the current study is inconsistent with the findings cited above. Though regular patterns may aid online processing, memory for individual words is instead facilitated by irregular patterns. We propose that this irregularity bias may reflect how uncommon metrically irregular speech is in English. Listeners might have an easier time recalling irregular speech due to more experience with it in their environment.

Implicit prosody and task-load interactions in native and late bilingual speakers of English

Elizabeth Pratt (The Graduate Center, City University of New York) & Eva M. Fernández
(Queens College, City University of New York)

Prosody—including implicit prosody—increases the saliency of grammatical markers and aids in the processing of structural relationships [1,2,3]. This project focuses on structural and prosodic effects during reading, examining their influence on agreement processing and comprehension in native (L1) and late bilingual (L2) English speakers. Our frame of reference is the Good-Enough Hypothesis [4,5], which proposes that comprehenders economize resources— particularly in demanding tasks—by using plausibility, pragmatics, and thematic templates to prioritize message extraction over complete syntactic analysis [4,5]. Our investigation integrates implicit prosody into this framework to clarify its role in processing demand in L1 and L2 populations.

We report data from a study that manipulated text presentation to influence implicit prosody, using sentences that induce subject-verb agreement attraction errors. Materials included 64 sets, crossing the factors of complexity (simple (1a)-embedded (1b)) and subject-verb agreement (grammatical-ungrammatical). Participants (n=23 L1, 16 L2) read items in one of three presentation formats: a) whole sentence, b) phrase-by-phrase (breaks indicated by |), or c) 500 ms word-by-word RSVP. Participants rated each sentence for grammaticality and responded to a comprehension probe.

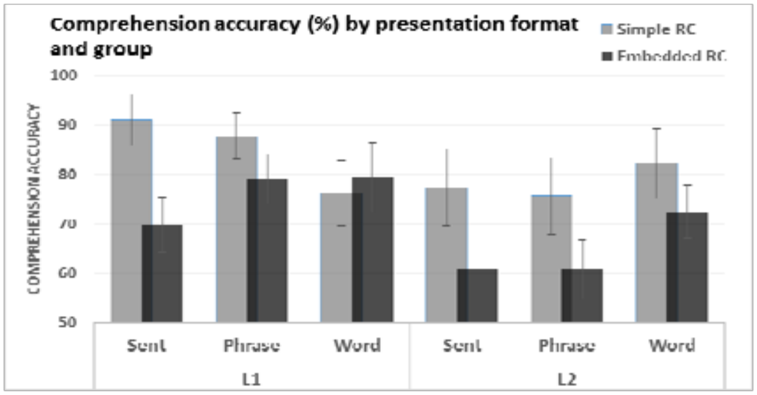
Following the Good-Enough Hypothesis, at baseline (whole sentence), comprehension accuracy should be lower for embedded materials than for simple, and the difficulty of the items should make error detection low overall. For L1ers, phrasal presentation should reduce this task load by facilitating implicit prosody, resulting in higher comprehension accuracy of the embedded materials, and higher error detection rates overall. L2ers, who are under greater task load due to lower reading fluency, may benefit from a slower word-by-word presentation.

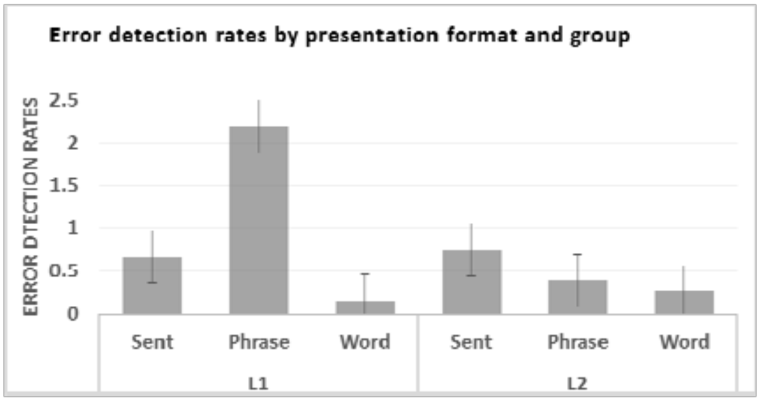
As anticipated, with whole-sentence presentation, comprehension accuracy was lower for embedded materials than for simple (p<.05), and overall error detection was low (grammatical vs. ungrammatical n.s., p>.20). For L1ers, these effects were mitigated with phrase-by-phrase presentation: comprehension accuracy was similar for embedded and simple materials (n.s., p>.10), and error detection rates were significantly higher than whole-sentence presentation. (p<.05). For L2ers, however, the effect on comprehension was mitigated with word-by-word presentation (n.s., p>.20), while error detection rates remained low in all conditions (p>.60).

Text presentation mediates processing load during reading, albeit differently in L1 and L2. For native speakers, phrasal presentation facilitates processing: prosody/syntax alignment augments comprehension of complex materials and improves overall syntactic processing. For L2 speakers, comprehension is improved with word-by-word presentation, where the benefits of slower input outweigh the cost of increased working memory load associated with this format.

Facilitating the projection of phrasal prosody onto text enhances performance for high-proficiency readers; for less proficient readers, task load is mediated by input speed. These differences provide insight into the interaction of cognitive task load and implicit prosody during both L1 and L2 reading.

1. a. The reporter | who called the senators every so often | write*/s awful stories for the paper.
   b. The reporter | who called the senators that Scott supported | write*/s awful stories for the paper.


Comprehension accuracy (%) by presentation format and group


Error detection rates by presentation format and group

References

[1] Fodor, J.D. 2002. Paper presented at Speech Prosody 2002.

[2] Frazier, L., Carlson, K., & Clifton, C. 2006. *Trends in Cognitive Sciences, 8*, 244–249.

[3] Kreiner, H. 2005. Paper presented at International Symposium on Discourse and Prosody.

[4] Ferreira, F. 2003. *Cognitive Psychology, 47*,164–203.

[5] Ferreira, F., Ferraro, V., & Bailey, K.G.D. 2002. *Current Directions in Psychological Science, 11*, 11–15.

Ellipsis incurs a penalty in parentheticals

Amanda Rysling, Charles Clifton, Jr., & Lyn Frazier (University of Massachusetts Amherst)

Multiple sources of evidence suggest that Not At Issue (NAI) and At Issue (AI) constituents are distinct. Formal semantics work has argued that interpretations of NAI and AI material are separately computed [4] (though see also [1]). Processing evidence indicates that NAI and AI material may utilize separate memory stores during sentence comprehension [3]. Despite apparent separateness, NAI and AI content interact [5]. Elided phrases in NAI constituents (e.g. parentheticals or appositives) may acceptably take antecedents in AI constituents [1]. Here we present evidence of a processing penalty for crossing the NAI-AI divide in ellipsis antecedent resolution.

Two auditory sentence naturalness-rating studies investigated the effect of AI antecedents for NAI ellipses. In the first experiment (n=48), NAI vs AI status of comment clauses was manipulated by using 'comma' intonation ([4], 'incidental' in [2]), to yield a parenthetical/appositive analysis of the material. Prosodically integrated AI clauses were contrasted with comma-intoned NAI clauses [examples (**1**) & (**2**) vs (**3**) & (**4**), where parentheses indicate comma intonation]. When the comment clause did not contain ellipsis, ratings were higher for NAI clauses than for AI ones [(**3**) > (**1**)]. Crucially, ellipsis resulted in a larger drop in acceptability in the NAI parenthetical condition than its AI prosodically integrated counterpart [(**3**)−(**4**) > (**1**)−(**2**)]. The second experiment (n=48) [(**3**) & (**4**) vs (**5**) & (**6**)] showed that the NAI parenthetical structure underlies the effect, not just the presence of a prosodic boundary separating an ellipsis site and its antecedent in entirely AI material.

Statistical analyses indicate that dispreference of ellipsis depends on whether it occurs in NAI or AI material. Mixed effects modeling with random slopes and intercepts by subjects and items revealed significant interactions of *ellipsis* and *information status* in both Experiment 1 (β=−0.53, t=2.62) and Experiment 2 (β=1.68, t=5.76). Experiment 1 shows that eliding material in an NAI parenthetical lowers acceptability more than eliding material in an AI comment. Experiment 2 shows that ellipsis across a simple prosodic boundary is not penalized.

The penalty observed for ellipsis in NAI structures might be due to the need to consult distinct memory stores for AI and NAI content [3]. If so, then a penalty for ellipsis should be present in a wide variety of structures, not just comment clauses, while a penalty for crossing the NAI-AI divide should be incurred by other dependencies, not just ellipsis. We are presently testing these predictions.

no ellipsis AI **(1)** [$_{\text{iP}}$A Frenchman I think it was a Frenchman] [$_{\text{iP}}$broke the record.]
ellipsis AI **(2)** [$_{\text{iP}}$A Frenchman I think it was] [$_{\text{iP}}$ broke the record.]
no ellipsis NAI **(3)** [$_{\text{iP}}$A Frenchman] [$_{\text{iP}}$ (I think it was a Frenchman)] [$_{\text{iP}}$ broke the record.]
ellipsis NAI **(4)** [$_{\text{iP}}$ A Frenchman] [$_{\text{iP}}$ (I think it was)] [$_{\text{iP}}$ broke the record.]
no ellipsis AI **(5)** [$_{\text{iP}}$ I thought it was a Frenchman,] [$_{\text{iP}}$ but it wasn't a Frenchman.]
ellipsis AI **(6)** [$_{\text{iP}}$ I thought it was a Frenchman,] [$_{\text{iP}}$ but it wasn't.]

Means and standard errors for all experiments, naturalness ratings 1 to 7 (7=high):

Exp 1	NAI (**3**, **4**)	AI (**1**, **2**)
−elide	4.34(0.14)	3.45(0.13)
+elide	3.61(0.13)	3.26(0.13)

Exp 2	NAI (**3**, **4**)	AI (**5**, **6**)
−elide	5.17(0.11)	5.60(0.09)
+elide	4.18(0.14)	6.29(0.08)

Selected References: [1]. Anderbois, Brasoveanu, & Henderson. (2015). *Journal of Semantics,* 32(1):93-138. [2]. Bonami & Goddard. (2007). *Proceedings of the 14th International Conference on HPSG*, pp. 25-45. [3]. Dillon, Clifton, & Frazier. (2014). *Language, Cognition and Neuroscience.* 29(4):483-498. [4]. Potts. (2005). *The Logic of Conventional Implicatures.* [5]. Syrett & Koev. (2014). *Journal of Semantics*. doi:10.1093/josffu07

Prosodic phrasing in German narrow focus constructions

Fabian Schubö (University of Stuttgart)

Background: This study tests for an impact of focus and givenness on the prosodic phrasing of sentences with clausal embedding in German. Prior research showed that givenness leads to prosodic reduction in form of post-focal deaccentuation and pre-focal pitch accent compression (e.g. Féry & Kügler 2008); It is unclear, however, if the presence of given elements also influences prosodic phrasing, in particular the formation of Intonation Phrases (IP). It has been found that the insertion of IP boundaries is triggered by the edges of syntactic clauses and that embedded clauses (but not root clauses) allow for variability in regard to the presence of a coinciding IP boundary (Downing 1970, Truckenbrodt, 2005). Clause boundaries and givenness may thus be hypothesized to constitute opposing forces in the formation of IP structure: Clause edges enforce the insertion of IP boundaries whereas given elements may prevent them.

Objective: The present study tested which of these opposing forces is overriding in German narrow focus constructions, i.e., whether the insertion of an IP boundary at a clause edge is prevented if it is followed or preceded by given material only. The prediction was that IP boundaries are absent when only given material follows or precedes. This prediction is based on the observations that pitch accents are absent on post-focal material and that given elements reject main stress (Féry & Samek-Lodovici 2006), both of which are conditions for IP formation.

Experiment: A production experiment was conducted which elicited sentences with an object clause (1) under three focus conditions: first, with broad focus, second, with a narrow focus on the object of the main clause (*Lehrer* 'teacher') and, third, with a narrow focus on the subject of the embedded clause (*Manuel*). The unfocussed material was explicitly given in a preceding context question. 18 items of this sort were recorded with six subjects (n=108). IP boundaries were detected by automated analysis of the f0 contour preceding the internal clause edge for every recorded utterance. The presence of an f0 rise from the object to the verb was taken as indicating the presence of a high boundary tone marking an IP edge.

Results and discussion: Speakers regularly inserted an IP boundary in the broad focus condition (87%), but not in the narrow focus conditions; a narrow focus in the main clause lead to IP-boundary insertion only in 2% of cases whereas a narrow focus in the embedded clause lead to IP boundary insertion in 35% of cases. The results support the hypothesis that givenness prevents the insertion of IP boundaries in constructions with clausal embedding, thus overriding the force of the internal clause edge to trigger an IP boundary. The difference between the two narrow focus conditions in regard to phrasing variability suggests that post-focal deaccentuation makes the insertion of a post-focal IP boundary impossible whereas main stress rejection of given elements reduces the likelihood of an IP boundary in pre-focal position, but does not rule out this possibility altogether.

(1)	*Cormelius*	*will*	*dem*	*Lehrer*	*melden,*	*dass*	*Manuel*	*eine*
	Cornelius	wants	the	teacher	report	that	Manuel	a
	Brille	*gestohlen*		*hat.*				
	glasses	stolen		has				

'Cornelius wants to report to the teacher that Manuel stole a pair of glasses.'

Downing, Bruce T. 1970. Syntactic structure and phonological phrasing in English. Doctoral dissertation, University of Texas at Austin. ◆ Féry, Caroline & Frank Kügler, 2008. Pitch accent scaling on given, new and focused constituents in German, *Journal of Phonetics* 36, 680-703. ◆ Féry, Caroline & Vieri Samek-Lodovici, 2006. Focus projection and prosodic prominence in nested foci. *Language* 82, 131-150. ◆ Truckenbrodt, Hubert. 2005. A short report on intonation phrase boundaries in German. *Linguistische Berichte* 203, 273-296.

Typologizing native language influence on second language intonation production: Three transfer phenomena in Japanese EFL learners

Aaron Albin (Indiana University - Bloomington)

Influence (or 'transfer') from one's first language (L1) when speaking a second language (L2) is a significant contributor to the pervasive cross-speaker variability in the linguistic use of F0. Three examples of ways one's L1 intonational phonology can affect L2 production are listed in (1). While there is now a substantial volume of research on the L2 acquisition of intonation (cf. Delais-Roussarie, Avanzi, & Herment (2015) and the studies cited therein), most studies have focused on a single transfer phenomenon in a single L1-L2 pairing. With the existing research, it is unclear whether all three of the transfer phenomena in (1) occur to the same extent or whether some are inherently more pervasive than others. To rectify this state of affairs, what is needed is a study collecting comparable data on how frequently multiple phenomena are attested in the same population of learners.

Toward this end, the present study analyzes data from the English Speech Database Read by Japanese Students, an 89,000+ soundfile speech corpus of elicited production from Japanese learners of English as a foreign language (EFL). The three cross-linguistic differences under examination as a source of transfer for this L1-L2 pairing are listed in (2), each exemplifying the corresponding phenomenon in (1). From the larger corpus, 14 sentences were selected for each of the three phenomena such that the relevant kind of transfer would be most likely to unambiguously appear, with approximately 22-28 learner tokens plus 11 native speaker tokens per sentence. Each token was segmented into the syllables and/or segments over the crucial region for each phenomenon. F0 tracks were stylized using a novel approach to quantifying not only F0 turning points but also the nonlinear shape of the transitions between neighboring turning points. The frequency of occurrence of each of the three transfer phenomena was determined by testing the concomitant empirical predictions about the shape of this stylized contour.

Results suggest that the three kinds of transfer phenomena under investigation do indeed occur at different frequencies: (a) the spurious insertion of an H- phrase-initially is vanishingly rare, (b) marking of pre-pausal boundaries with L% is incredibly frequent (substantially more so than native speakers), and (c) beginning boundary rises inside the utterance-final syllable occurs at an intermediate frequency. Interestingly, of the three, only (c) - boundary rise alignment - appears to be tied to proficiency (being significantly less common in advanced learners).

It is argued that this particular hierarchy is obtained because these three different kinds of transfer represent qualitatively different ways that the intonational phonology of two languages can mismatch. The results are discussed in terms of a proposal for a broader "typology of intonational transfer" that attempts to sketch out the space of possible ways one's L1 intonation system can influence L2 production, thus establishing a framework that can be applied in future research. By advancing our understanding of exactly how and why L2 learners diverge from native norms, the present study sheds light on this important but ill-understood facet of prosodic variability.

Delais-Roussarie, E., Avanzi, M., & Herment, S. (2015). *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations* New York: Springer.

(1) a. A learner may produce a tone in a position that is phonologically empty in the L2 but where a tone is obligatory in the L1.
b. A learner may parse their utterance in the L2 into too many or too few prosodic phrases due to differences in the number of pitch accents per phrase in the L1 and the L2.
c. A learner may phonetically realize an L2 tone in ways corresponding to an analogous L1 tone due to the superficial similarity between the two.

(2) a. Japanese marks the left edge of accentual phrases with a phrasal H- on the second mora, whereas prosodic phrases in English generally have no tone before the first pitch accent.
b. Japanese accentual phrases can only contain one H*+L pitch accent, whereas English prosodic phrases can contain multiple pitch accents in sequence.
c. A boundary rise in Japanese must begin at the utterance-final syllable, whereas boundary rises in English often begin at the nuclear stressed syllable (which can be several syllables earlier).

Some clustering with k-means: Acoustics of a prosodic contrast in a corpus of spontaneous speech

Anca Chereches (Cornell University)

Many theories of prosody (e.g. Selkirk 1996, Calhoun 2006) predict that function words like *some* will be highly reduced in English (1a), except in a few specific environments, such as under focus (1b-c). The present study quantifies the link between focus and the acoustic realization of the quantifier *some* using a novel technique.

1. a) **[sm]** people brought us fresh cilantro.
   b) Cilantro? **[sʌm]** people love it, **[sʌm]** people hate it.
   c) I guess **SOME** people like cilantro.

A previous corpus study linked focus to prosodic prominence (Calhoun 2006), but it did not zoom in on function words and it was conservative in its focus annotation. The annotation guidelines favor explicit contrasts (1b), and recommend avoiding trickier cases of implicit contrast, such as in (1c), where a prominent *some* can trigger the scalar inference that not all people like cilantro. Such inferences are assumed by most semanticists to be very common, but judgments are actually sensitive to contextual information, including (potentially subtle) prosodic cues (de Marneffe & Tonhauser 2014). Other types of function words (prepositions, conjunctions, etc.) get focused more rarely, and often to correct some previous expression. Corrective focus tends to be very marked acoustically. Thus, *some* provides an interesting case study, where a word is predicted by prosodic theories to shift from low to high relative prominence under focus in general, but for types of focus like implicit contrast, it is not yet clear how this is reflected in its acoustics.

To address this question, I collected from online sports radio talk shows a corpus of utterances of *some* followed by *people* (resulting in 143 focused and 54 unfocused *some*s) and by *money* (4 focused, 194 unfocused *some*s), for a total of 395 tokens. I measured vowel duration, intensity, formants, F0 range, and size and alignment of F0 extrema. I manually annotated for focus. Since I felt relatively uncertain in labeling implicit contrast (e.g. scalar inferences), I did not train a supervised model like logistic regression. Instead, I ran k-means clustering on all combinations of acoustic measurements (8191 combinations) and selected for analysis the 177 models which best matched my manual labels and performed better than the highest baseline (85.3%).

The sheer number of successful unsupervised models suggests that for this dataset, focus correlates with a specific acoustic realization of *some*, which can best be characterized by increased duration, more extreme formants and an F0 peak (all measurements taken from the vowel in *some*). Many combinations of features achieve identical or near-identical cluster separation, suggesting that there is a great deal of redundancy in acoustic cues to focus. The single feature that appears in all winning models is duration of the *some* vowel. On its own, it separates the tokens with 86.7% accuracy, compared to the highest accuracy of 88.5% for duration, combined with formant and F0 values. Preliminary error analysis suggests that misclassified tokens were also difficult to label. In future work, we intend to quantify this labeling uncertainty by increasing the number of annotators.

To conclude, the present study supports prosodic analyses of function words which account for the strong influence of focus. Despite uncertainty in labeling focus when the contrast was implicit, acoustically we found remarkably good separation between focused and unfocused *some*s, especially along the duration dimension, and to a smaller extent formants and F0. Future work can strengthen these results by recruiting extra annotators and further processing the dataset to allow for more sophisticated normalization techniques.

References

- Calhoun, Sasha. 2006. “Information Structure and the Prosodic Structure of English: A Probabilistic Relationship.” Ph.D., University of Edinburgh.
- de Marneffe, Marie-Catherine, and Judith Tonhauser. 2014. “Prosody Affects Scalar Implicature Generation.” Poster presented at CUNY 2014.
- Selkirk, Elisabeth. 1996. “The Prosodic Structure of Function Words.” Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition 187: 214.

Interpreting patterns of variability in the realization of English intonation contours

Jonathan Barnes[1], Nanette Veilleux[2], Alejna Brugos[1], & Stefanie Shattuck-Hufnagel[3]
[1] Boston University, [2] Simmons College, [3] MIT

Recent work on the phonetic realization of intonational pitch targets has made it clear that this mapping must take account of meaningfully-patterned variability substantially more extensive and complex than previously understood. Where Autosegmental-Metrical approaches once described phonetic realization as “a matter of reaching a certain pitch level at a particular point in time” (4), perception and production experiments have since demonstrated the importance of F0 contour shape features not easily reducible to the location in time or F0-space of any single point or points along the contour (2, 5, 8, 9).

Models such as Tonal Center-of-Gravity (TCoG, 2) take this to mean that instead of seeking out turning points, velocity maxima, etc., listeners make holistic judgements about the overall F0 distribution within some salient region of the utterance. TCoG integrates F0 samples with a weighted average, yielding a single position in time and F0-space representing the F0-event’s “center-of-gravity”. But how do listeners accomplish this integration: Do all F0 samples within the region of interest exert equal influence on perception, or do some weigh more than others? Recent evidence suggests that samples from lower sonority segments (6, 7) contribute less to decisions regarding F0-event timing (1) and scaling (3) than do higher sonority samples. Superficially identical F0 contours might thus be interpreted differently by listeners, depending on the segments involved. This study delves further into the interaction between tones and their segmental hosts.

First, we created a series of five differently-shaped rise-fall-rise contours realized over the same base sentence: a linear rise followed by a linear, concave or convex fall, and a linear fall preceded by a concave or convex rise (Fig. 1). The timing of each was manipulated to create two distinct continua: 1) seven steps from early rise-fall (H+!H* L-H%) to medial rise-fall (L+H* L-H%) and 2) seven steps from medial L+H* to late-timed L*+H. 19 subjects heard two pairs of contours (AXBX: A and B represent endpoints of a timing continuum, X represents a contour from that continuum) and were asked which contour pair “matched”. The TCoG model predicts that contour shape, alongside timing, should influence listener categorizations, with domed rises and scooped falls biasing listeners toward “earlier” responses, and scooped rises and domed falls biasing toward “later”.

For the early-to-mid timing continuum, this indeed happened. In the mid-to-late continuum, however, although rise curvature was influential, fall curvature had little effect (Fig. 2). The reason is that in the early-to-mid continuum, both the rise and fall are realized primarily over relatively high intensity vowels, allowing them to contribute about equally to the TCoG location. For mid-to-late, however, the fall coincides largely with two sonorant consonants and a short reduced schwa, where F0 samples have less impact. Figure 3 shows how a TCoG model assigning lower weights to F0 samples from less sonorous segments accounts for subjects’ response patterns. This result underscores the danger of analyzing F0 patterns without considering the segmental material whose periodicity gives rise to them.

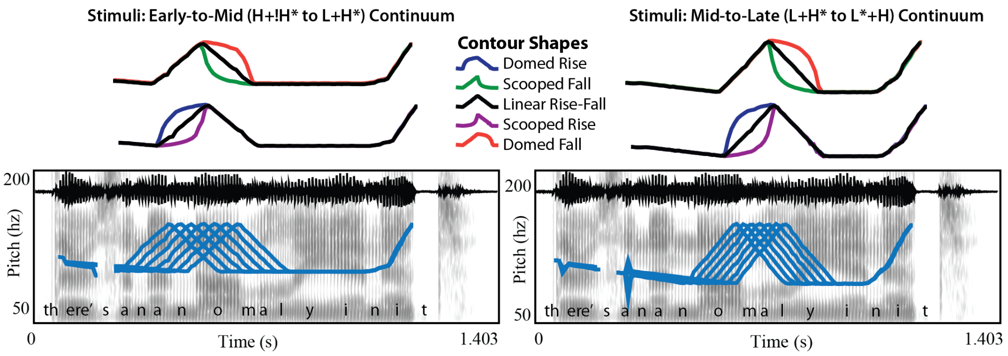


Figure 1. Shapes of accent-related rises/falls. Examples of rise-fall timing continua from early (H+!H*) to mid (L+H*) and mid to late (L*+H) on frame sentence *'There's an anomaly in it'.*

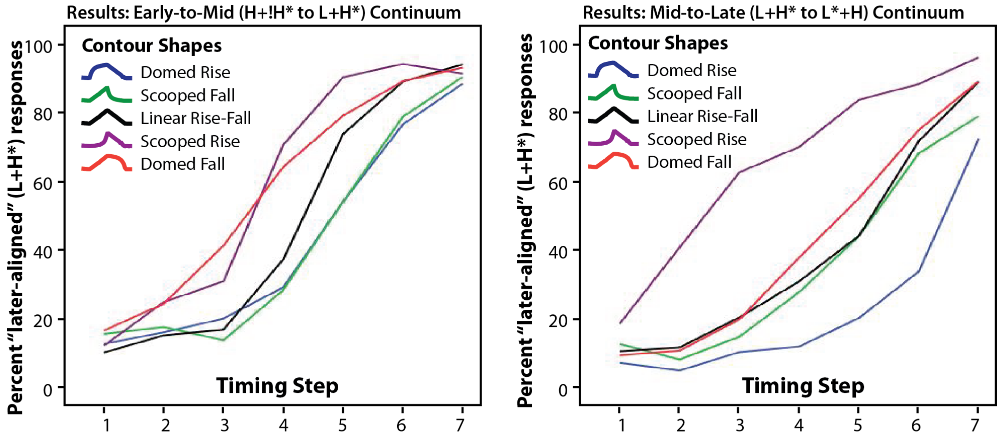


Figure 2. (Left) Early-to-mid continuum: scooped rises/domed falls bias listeners toward L+H*; domed rises/scooped falls do the opposite. (Right) Mid-to-late continuum: scooped rise biases listeners toward L*+H; domed rise toward L+H*; scooped and domed falls similar to linear.

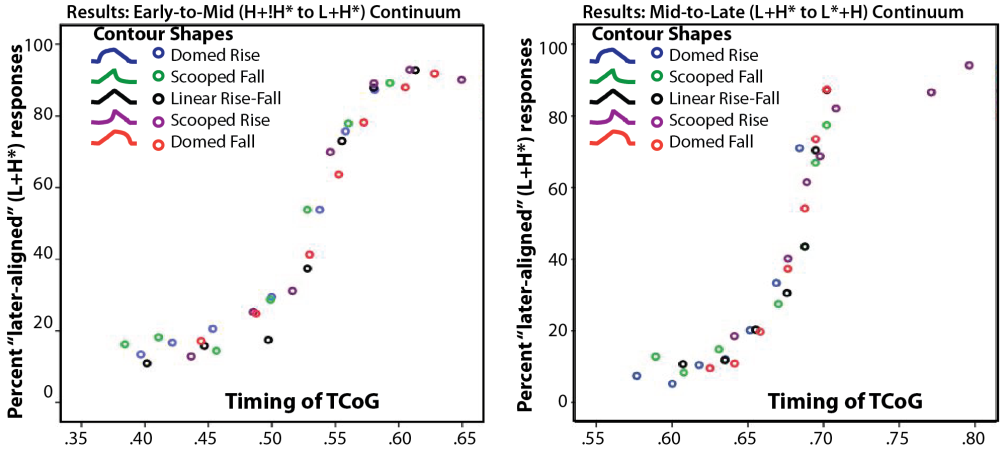


Figure 3. Location of TCoG in time using lower weights for lower sonority segments yields tightly-clustered s-shaped responses. Contour shapes fit the same curve as a function of their TCoG.

References
[1] Barnes et al. (2012). In: O. Niebuhr (ed.), Prosodies –Context, Function, Communication, 93-118. Berlin/New York: de Gruyter. [2] Barnes et al (2012). Laboratory Phonology 3(2), 337-383. [3] Barnes et al (2014) Speech Prosody 7, 1125-1129. [4] Bruce, G. (1977) Travaux de l'Institut de Linguistique de Lund 12. Lund, Sweden: CWK Gleerup. [5] D'Imperio, M. (2000) Thesis OSU [6] Flemming, E. (2008) In M. Embarki and C. Dodane (eds.), La Coarticulation: Indices, Direction et Representation. [7] Gordon (2001) Studies in Language 25: 405–444. [8] Knight. R. (2008) Lang. & Speech 51, 3: 223-244. [9] Niebuhr, O. (2007) Phonetica 64(2) 174-193.

Perceptual learning of intonation contour categories: The role of utterance duration

Vsevolod Kapatsinski & Paul Olejarczuk (University of Oregon)

In recent work, we extended work on perceptual category learning (Gibson & Gibson 1955, Posner & Keele 1968) to the acquisition of intonation contour categories (Anonymous, in revision). We used a flat prototype (-----), a final-fall prototype (˜˜˜˜\), and a two-peaked prototype (/_/_) and created distortions of these prototypes to model within-category variability. We presented adult and 9- to 11-year-old native English speakers with examples of each contour category. Training examples were minor perturbations of the prototype, averaging out to the prototype. This created well-separated categories with multiple partially redundant and acoustically variable but individually necessary features. We suggest that this category structure is characteristic of intonation contour categories, as it is characteristic of other categories of linguistic forms (Kapatsinski 2014).

We hypothesized that an adult would require all of the necessary features of a contour category to be present before classifying a novel contour into the category, the same way that an adult requires all phonological features of a word like [blæk] to have been intended by the speaker to perceive it as that word. A one-feature intentional deviation, as in [blæg], is big enough to block categorization into the same category, e.g. eliminating repetition priming (Stockall & Marantz 2006; Darcy et al. 2009). Despite the fact that [blæg] is more similar to [blæk] than to any other word, and the features of [blæ[Dorsal]] are sufficient for rejecting all words that compete for recognition with [blæk], [blæg] is not perceived as a realization of /blæk/ without contextual evidence that a voiceless /k/ was actually intended or retraining of the phoneme-sound mappings. The voicelessness of [k] in [blæk] is thus, sans perceptual retraining, a necessary feature. Similarly, while the contour /_/_ can be distinguished from all other contours in the experiment by, say, the initial rise, we hypothesized that adults would require all features of /_/_ to be present, turning characteristic features into necessary ones. However, paying attention to all of the features of a contour presents a working memory challenge. Children, whose working memory capacity is not yet mature (Luna et al. 2004), may not be able to keep track of all contour features, thus requiring only some of them to be present (Ward & Scott 1987, Thompson 1994).

On test trials, participants heard exemplar contours and categorized them as something said by an experienced creature or by a novel creature they had not encountered. Choosing the novel creatures was considered rejection, similar to perceiving [blæg] as a pseudoword. As expected, adults made more rejections, particularly judging that /____ could not have been said by the creature that said /_/_, whereas children accepted /___. However, surprisingly, both children and adults accepted __/_. For the present experiment, we reinstantiated the contours over shorter syllable sequences (7 rather than 16 syllables). With these less memory-demanding stimuli, adults rejected ___/_. These results support attributing the children's high tolerance of deviation from prior experience to memory demands posed by keeping track of all of the necessary features of a temporally extended contour.

Darcy, I., et al. (2009). Phonological knowledge in compensation for native and non-native assimilation. F. Kügler et al. (Eds.), *Variation and gradience in phonetics and phonology*. Mouton.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32-41.

Kapatsinski, V. (2014). What is grammar like? A usage-based constructionist perspective. *Linguistic Issues in Language Technology, 11,* 1-41.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*, 1357-1372.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *JEP*, *77*, 353-363.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, *1*(1), 85-123.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development, 65*, 1627-1645.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family resemblance concepts. *Memory & Cognition, 15,* 45-52.

Double focus of adjacent words in yes/no questions

Jonathan Howell (Montclair State University)

The Problem: Given a sequence of syntactic constituents *A B*, formal semantic theories of focus (e.g. Rooth 1992) predict four possible configurations:

(1) *broad* $[\text{ A B }]_F$ *early* $[\text{ A }]_F$ B *late* A $[\text{ B }]_F$ *double* $[\text{A}]_F$ $[\text{B}]_F$

Models of prosodic prominence may either have a corresponding set of four patterns or else a non-direct mapping in which the one prosodic pattern corresponds to two or more focus configurations. Strictly syntagmatic models of prosodic prominence, such as those found in the early metrical phonology literature, allow for only two patterns of prominence: weak-STRONG and STRONG-weak (cf. Ladd 1991).

However, models which adopt paradigmatic prominence (e.g. primary vs. secondary stress; or different pitch accent types) in addition to syntagmatic prominence distinguish four different prosodic categories corresponding to the different focus configurations. Autosegmental-metrical phonology accommodates this by positing a hierarchy of prosodic categories: prominence remains syntagmatic at each level of the hierarchy. For example, the double focus pattern may correspond to two adjacent prosodic phrases and adjacent nuclear accents (see for example Ladd 2008). This analysis predicts a significant difference in production for all four focus configurations.

The considerable literature on focus projection (see Selkirk 1984; Cinque 1993; Zubizarreta 1998; among others) conflates *broad* and *late* focus to a single weak-STRONG pattern of prominence. This analysis predicts no significant difference in production for broad and late focus.

Previous studies examining double focus have not examined adjacent foci. Eady *et al.* (1986) found significant differences in duration and mean *F0* across the four focus conditions in declarative sentences. Welby (2003) and Jannedy (2002) found that listeners tended to associate particular ToBI-annotated declarative utterances with particular focus conditions. It is unclear, however, how these results will generalize (i) to adjacent foci and (ii) to non-declarative contexts.

Experiment: 53 native English speakers were recorded reading 12 target utterances: yes-no questions. Each utterance was presented following one of four focus contexts, in a Latin-square design (cf. 1). The target words were verb-noun sequences consisting only of sonorants, in order to mitigate *F0*-tracking errors. Data were annotated by forced-alignment (Gorman *et al* 2011.) and acoustic measures were extracted using Praat.

We used linear residualisation (Breen et al. 2010) to remove acoustic variation due to speaker and item. A discriminant function model for focus condition using word duration and mean *F0* for the target words (e.g. *email* and *Owen*) was significant (Wilks 0.96634; F=3.1307; p<.01).

The results demonstrate that speakers *can* produce four distinct prosodic patterns corresponding to the four different focus configurations. However, in a leave-one-out LDA classification, the overall prediction of the model was only 28%. This result suggests that there is no categorical production difference; rather, we hypothesize that acoustic cues are available to optionally disambiguate between focus categories.

Finally, these results do not support uniquely syntagmatic models of prosodic prominence (according to which a word is simply prominent or not prominent); and they do not support theories of focus projection, which predict no difference in production between broad and late focus.

(2) **Target:** Did you also email Owen last night?
Early: I heard you got drunk and texted Owen last night.
Late: I heard you got drunk and emailed a bunch of your work friends last night.
Double: You said you already called Ron and you texted Craig.
Broad: I heard you found time to make popcorn and veg out.

Production data is publicly available at: http://dx.doi.org/10.7910/DVN/29343

References: **Cinque**, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry 24*:239-297. **Eady**, S. J. and W. E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America 80*: 402–415. **Gorman**, Kyle, Jonathan Howell and Michael Wagner. 2011. Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. *Canadian Acoustics*. 39.3. 192–193. **Jannedy**, Stefanie. *Hat Patterns and Double Peaks: The Phonetics and Psycholinguistics of Broad versus Late Narrow versus Double Focus Intonations*. 2002. Doctoral dissertation. **Ladd**, D. Robert. 1991. One word's strength is another word's weakness: Integrating syntagmatic and paradigmatic aspects of stress. *Proceedings of the Seventh Eastern States Conference on Linguistics*. **Selkirk**, Elizabeth O. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, M.A.:MIT. **Welby**, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *Language and Speech 46*:53-81. **Zubizarreta**, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Cambridge, Mass..: MIT Press.

****

Natural speech perception by L1 and L2 speakers of English

Shinobu Mizuguchi[1], Jennifer Cole[2], Gabor Pinter[1], Koichi Tateishi[3], & Tim Mahrt[2]
[1] Kobe University, [2] University of Illinois at Urbana-Champaign, [3] Kobe College

Keywords: natural speech, perception, L2 speakers, prominence, boundary

1. Experiment

This study investigates how prosodic boundaries and prominent words are perceived in spontaneous English speech by speakers in three different language groups: native speakers of English (NS), intermediate level Japanese EFL learners (Int), and advanced Japanese EFL learners (Adv). The experiment paradigm (Rapid Prosody Transcription, RPT) and the data from 16 native speakers were taken from prior studies [1],[2]. The non-native data was collected by rerunning a subset of the RPT tasks from the native-targeted experiments, which used excerpts from the Buckeye Corpus, with 108 intermediate and 15 advanced level Japanese undergraduate students.

2. Results and Discussion

Listeners marked the location of perceived prominence and boundaries, and for each word an average prominence score (p-score) and boundary score (b-score) were calculated, with values ranging from zero to one. When all listeners agree on the rating of a given word it will have a score of zero (nobody marked it as being prominent or final in a phrase) or one (everyone marked it). Fig.1 summarizes the inter-listener agreement in the three groups for the location of prosodic boundaries and prominences over the 11 audio stimuli. As expected from the previous studies, there was greater agreement on boundaries than prominence in all groups; also, native speakers exhibited greater inter-listener agreement than EFL learners, and the Adv learners exhibited greater inter-listener agreement than the Int learners. A possible interpretation of this is that advanced learners have acquired more native-like behavior.

2.1 Correlation

Kendall's test was used between the pairs NS-Int and NS-Adv to see if there is a stronger correlation of p-scores and b-scores between the NS and Adv group. Fig. 2 displays NS-Int correlation coefficients plotted against NS-Adv coefficients, calculated over each utterance. The distribution of data points in the p-score plot implies that the advanced group achieved stronger correlation with native speakers in prominence. No such tendency is observed for boundary perception. Although agreement about boundaries increased with proficiency, it did not approximate native speaker behavior.

2.2 Role of Syntax

The previous study [2] shows that syntax directly influences boundary perception independent of the acoustic evidence. Fig.3 compares the number of mean boundary and perception markings, and shows that advanced learners are closer to native speakers in prosody perception marking than in boundary perception marking. Fig.4 shows the frequency of syntactic cues used at the left edge, where prosodic phrases start, and at the right edge, where prosodic phrases end. The NS group mainly uses CPs at the right edge and Ns at the left edge. Although the Adv group shows the same tendency, a crucial difference is found in the boundary perception of D(iscourse) M(arker)s, Conjunctions and Adverbs. DMs tend to be inserted between major phrases as markers of breaks, but EFL learners, regardless of levels, fail to perceive them.

3. Conclusion

This paper argues (i) JEFL learners improve in perception as they advance in their proficiency, and (ii) they are poor at using syntactic boundary cues of DMs, Conjunctions and Adverbs. To scrutinize the correlation between acoustic and syntactic cues is our next task.

References

[1] Y-S. Mo, J. Cole and E-K. Lee. 2008. 'Naïve listeners' prominence and boundary perception', in *Proceedings of the 4th International Conference of Speech Prosody*, pp.733-738.
[2] J. Cole, Y-S. Mo and S-D. Beak. 2011. 'The role of syntactic in guiding prosody perception with ordinary listeners and everyday speech', *Language and Cognitive Process*, Vol. 25, pp.1141-1177.

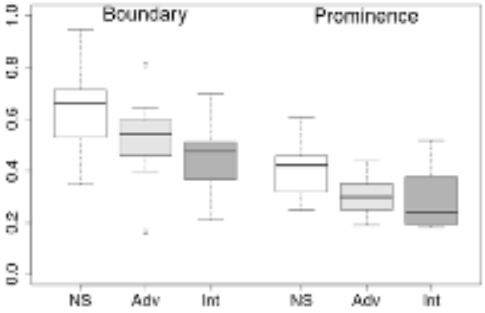


Fig. 1: Inter-speaker agreement (Fleiss' kappa)

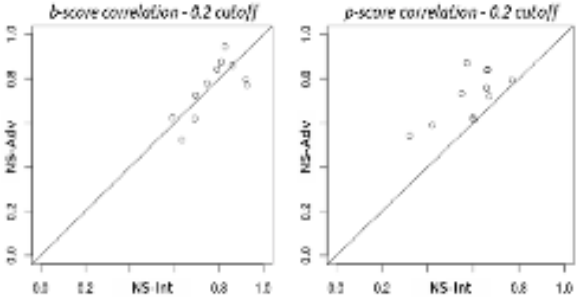


Fig. 2: p-score & b-score correlation (p-score/b-score > 0.5)

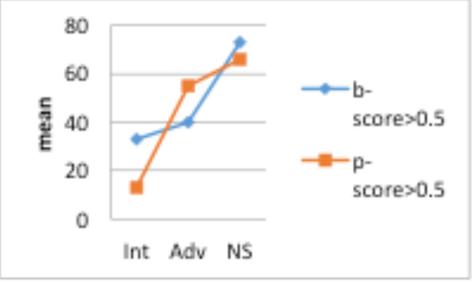


Fig. 3: Number of markings (mean)

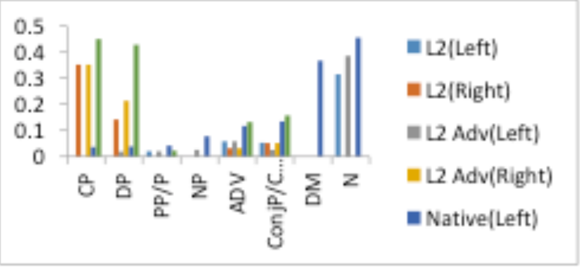


Fig. 4: Frequency of boundary perception (b-score>0.5)

Accent phrases duration constraints

Phillipe Martin (Université Paris Diderot)

As a language without lexical stress, French provides interesting examples to investigate prosodic variability, and in particular variability in phrasing. Indeed in French, accent phrases (AP), defined by prosodic events aligned on their right boundaries, can contain more than one lexical words (Noun, Adjective, Adverb or Verb) or even only grammatical words (Pronoun, Conjunction...). In a sentence such as *Max adore la région de Meaux mais pas que* “Max loves the region of Meaux but not only”, successive AP could be [*Max adore*] [*la région de Meaux*] [*mais pas que*] or [*Max*] [*adore*] [*la région*] [*de Meaux*] [*mais pas que*], depending on the speech rate, the last AP containing only grammatical words (stressed syllables are underlined).

Meigret (1550!) noticed that long words of more than seven syllables, normally stressed on their last syllable, could not be pronounced without at least one more (secondary?) stress located at some internal morphological boundary. For example *anticonstitutionnellement* (9 syllables) “unconstitutionally” and *paraskevidekatriaphobie* (10 syllables) “fear of Friday 13”, have to be pronounced with more than one single stressed syllable, as *anticonstitutionnellement* and *paraskevidekatriaphobie*, therefore dividing each word into two or three AP’s, *anti*, *constitutionnellement*, *paraskevi* “Friday”, *dekatria* “Thirteen” and *phobie* “Fear”.

Analysis of rather large spontaneous speech corpora reveals that this limitation does not pertain to the number of syllables but rather to the time it takes to pronounce them, the limit for the longest AP being in the order of 1200 ms. Furthermore, if a large number of syllables gets packed into a single AP, their duration is reduced so that the average syllabic duration in an AP varies with the number of syllables they contain, between 250 ms (one stressed syllable AP) to 1200 ms for 8 syllables (Martin, 2014).

Looking at brain waves properties, it is intriguing to note that the range of duration of Delta waves, from 250 ms to 1200 ms, corresponds closely to the range of AP duration, whereas the range of Theta brain waves, from 100 ms to 250 ms, correspond to the range of syllabic duration. As Theta waves optimize the perception of syllables (Henry & al., 2012), and given that Delta waves do synchronize Theta waves (Ghitza & al., 2013), this observation may provide an explanation for the necessity to stress syllables periodically, at least every 1200 ms or so, in order to synchronize the perception of syllables.

The correspondence between AP and Delta waves duration range may also explain eurhythmicity observed in French (Wioland, 1985), aiming to balance the successive IP duration, either by realizing IP’s possibly non-congruent with syntax by keeping their number of syllables similar, or by keeping the syntactic congruence by modulating the duration of AP to achieve eurhythmicity.

References

Oded Ghitza1, Anne-Lise Giraud & David Poeppel (2013) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence, *Frontiers in Human Neuroscience* www.frontiersin.org, January2013, Volume6, Article 340.

Henry, Molly & Jonas Obleser (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior, PNAS, www.pnas.org/content/early/2012/11/.../1213390109

Martin, Philippe (2014) Spontaneous speech corpus data validates prosodic constraints, *Proceedings of the 6th conference on speech prosody,* Campbell, Gibbon, and Hirst (eds.), 525-529.

Meigret, Louis (1550) *Le tretté de la grammére françoéze*, Paris, C. Wechel. http://gallica.bnf.fr/ark:/12148/bpt6k507854/f1.image

Wioland François (1985) *Les structures rythmiques du français*, Slatkine-Champion, Paris.

Key words: accent phrase, phrasing, delta waves, theta waves, eurhythmicity

Prosodic disambiguation of conditional vs. logical conjunction

Joseph Tyler (Morehead State University) & Ezra Keshet (University of Michigan)

Early work on prosodic disambiguation argued that ambiguous sentences could be disambiguated by prosody when the different meanings corresponded to different groupings of "contiguous constituents" (Lieberman, 1967, p. 110). Experimental work has supported this account (Lehiste, Olive, & Streeter, 1976), and gone on to explore how prosodic disambiguation of bracketing contrasts works (Price, Ostendorf, Shattuck- Hufnagel, & Fong, 1991; Snedeker & Trueswell, 2003). Here we present a series of studies testing speakers' and listeners' ability to prosodically disambiguate conditional from logical conjunctions, an ambiguity defined not by different bracketings but by semantic relationships between clauses.

Scholars have noted that and-conjoined clauses like (1) can be ambiguous between conditional conjunction (2) and logical conjunction (3) (Culicover & Jackendoff, 1997). And while prosody has been argued to disambiguate between CC and LC meanings (Pierrehumbert & Hirschberg, 1990), such claims are based on impressionistic judgments and have not been tested experimentally.

In a production study, 24 participants were recorded in a sound proof booth reading aloud 22 sentences like (1), once for CC and once for LC. The resulting sentences' clauses, conjunctions and inter-clausal silences (if present) were then queried with a Praat script for duration, pitch, and intensity information. Results show CCs had longer clauses but were shorter between clauses (fewer silences and shorter conjunctions). CCs were produced significantly quieter in clause 2, and on the conjunction. CCs also had lower f0max and initial pitch in both clauses and the conjunction, but higher final pitch on clause 1.

A perception study with fully-crossed design between speaker, item, and speaker's intended meaning (CC vs. LC), resulting in 44 different blocks of stimuli, tested whether listeners could retrieve speakers' intended meaning. Of 100 total participants, with at least two per block, results using metalinguistic glosses like (2) and (3) as a dependent variable showed no effect of speaker intended meaning on listener interpretation. While there was no overall disambiguation effect, one speaker was much more successful than all others (66% match rate). A second perception study using just this speaker's productions showed a highly significant effect of intended meaning on listener interpretation (t=9.31). While clause 1 final pitch seems involved, the most relevant contrast was an accented conjunction biasing towards LC.

Given prior work claiming Focus structures disambiguate CC from LC (Keshet, 2013) and the results above for clause 1 final pitch, another study tested the role of a rise-fall-rise contour on clause 1 to bias towards CC. The first author produced sentences like (1) with a rise-fall-rise contrastive contour (L-H* L-H%) on clause 1 to bias towards CC and a simple fall (H* L-L%) on clause 1 to bias towards LC. Results show listeners were above chance in retrieving the speaker's intended meaning (t=3.38).

The results demonstrate prosody can disambiguate CC/LC ambiguities, with an accented conjunction biasing towards LC and a rise-fall-rise contour biasing towards CC. Prosody can disambiguate more than syntactic bracketing contrasts.

(1) April brings her beagle, and everyone else stays home.
(2) If April brings her beagle, then everyone else stays home. (CC)
(3) Two things are true: April brings her beagle, and everyone else stays home. (LC)

Culicover, Peter W., & Jackendoff, R. (1997). Semantic subordination despite syntactic coordination. *Linguistic Inquiry, 28*, 195–217.

Keshet, Ezra. (2013). Focus on Conditional Conjunction. *Journal of Semantics, 30*(2), 211-256. doi: 10.1093/jos/ffs011

Lehiste, I., Olive, J., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. [10.1121/1.381180]. *J. Acoust. Soc. Am., 60*(5), 1199.

Lieberman, Philip. (1967). *Intonation, perception, and language*. Cambridge: M.I.T. Press.

Pierrehumbert, Janet, & Hirschberg, Julia. (1990). The Meaning of Intonation in the Interpretation of Discourse. In P. Cohen, J. Morgan & M. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge MA: MIT Press.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*(6), 2956-2970.

Snedeker, Jesse, & Trueswell, John. (2003). Using Prosody to Avoid Ambiguity: Effects of Speaker Awareness and Referential Context. *Journal of Memory and Language, 48*(1), 103-130.

Perceived prosodic variability predicts gesture rates in spontaneous speech

Aisha J. Bhutto & Maureen Gillespie (University of New Hampshire)

Extra-linguistic aspects of language production such as intonation patterns (e.g., Gaudio, 1994), speech rate (e.g., Miller et al., 1984), and gesture rates (e.g., Gillespie et al., 2014) vary from speaker to speaker and across situations. Models of language production do not fully account for these sources of variability (e.g., Bock & Levelt, 1994); however, some attempts have been made to incorporate gesture (e.g., Krauss et al., 2000) and the creation of prosodic structure (e.g., Watson et al., 2006) into these models. Variability in both prosodic patterns and gesture rates has been tied to planning processes in language production (e.g., Gillespie et al., 2014; Watson et al., 2006). The current exploratory work examined whether perceived individual variability in prosodic patterns is predictive of gesture rates in spontaneous speech.

In the current study, 37 participants watched 30-second video clips of speakers giving presentations (obtained from YouTube lectures) with the accompanying audio (Vid), and 31 participants listened to the audio alone (Aud). All participants then rated the speaker's pitch variability, volume variability, speech rate, and gesture rate on a 0 to 100 scale. Crucially, the participants in the Audio Only version could only guess gesture rates, as they did not see the speaker.

Measures of prosodic variability for each speaker were highly correlated in the both versions of the task (Volume: r = 0.64 Pitch: r = 0.75, Rate: r =0.74; ps < .001), suggesting that judgments of prosodic properties were not influenced by viewing the speaker. Additionally, perceived (Vid) and guessed (Aud) gesture rates were significantly correlated (r = .38, p < .02). In both versions of the task, increased perceived pitch variability (Vid: t = 4.49; Aud: t = 5.94) and faster speech rates (Vid: t = 3.27; Aud: t = 3.12) were associated with higher gesture rates, while perceived volume variability did not predict gesture rates (ts < 1.7). Objective measures of prosodic variability and gesture rates are currently being obtained and resulting patterns will be discussed in relation to the perception study.

Interestingly, pitch variability and speech rate were identified as being predictive of gesture rate whether or not the raters could see the speaker, suggesting that listeners are able to pick up on some speech properties that are associated with gesturing. While this study was exploratory in nature, it does hint at the possibility that these sources of extra-linguistic variability may be linked in the language production process. Current work in the lab examines this possibility and preliminary results will be discussed.

Bock, J.K. & Levelt, W.J.M. (1994). *Handbook of Psycholinguistics.* (p. 945-984).
Gaudio, R.P. (1994). *American Speech, 69,* 30-57.
Gillespie, M., James, A.N., Federmeier, K.D., & Watson, D.G. (2014). *Cognition*, *132,* 174-180.
Miller, J.L, Grosjean, F., & Lomanto, C. (1984). *Phonetica, 41,* 215-225.
Watson, D.G., Breen, M., & Gibson, E. (2006). *JEP:LMC, 32,* 1045-1056.

Prosodic strategies of L1 and L2 speakers for attitudinal expressivity in USA English

Donna Erickson[1], Albert Rilliard[2], Takaaki Shochi[3], & João Antonio de Moraes[4]
[1] Kanazawa Medical University, Japan, [2] LIMSI-CNRS, France, [3] CLLE-ERSSaB UMR 5263, France,
[4] Laboratorio de Fonetica Acustica, FL/UFRJ/CNPq, Brazil
rilliard@limsi.fr, ericksondonna2000@gmail.com, shochi38@gmail.com, jamoraes3@gmail.com

Prosody changes with the speaker's communication goals, the context of interaction, and the cultural origin of the speaker (Wichmann, 2002). Various attitudinal expressions have been described for different languages in the literature (Uldall, 1960; Martins-Baltar, 1977; Fujisaki & Hirose, 1993; de Moraes, 2008; Shochi et al., 2009). The labels used to describe them vary from one language to another, as do the contexts of similar labels, rendering difficult the comparison of prosodic differences. We describe here a recording paradigm that was developed in order to bypass this limitation of translation of labels, with the aim of comparing prosodic strategies of speakers of different cultural origins.

The present study focuses on USA English attitudes, as produced by both USA English speakers (L1) and L2 English speakers (from Japan and France). L1 listeners rated the quality of the audio-visual performances; a subset of the best expressions was then used for which a different set of L1 listeners were asked to recognize 9 attitudes, in both audio and visual modalities. Results show that the cultural origin of the speaker affects the mean recognition performance, but only marginally the categorization of the expressions.

To better understand the (dis)similarity across cultures, fundamental frequency and intensity of the stimuli were analyzed. The observed distribution is discussed in light of two theoretical propositions: the Frequency code (Ohala, 1994) and the Effort code (Gussenhoven, 2004). The observations suggest an overall coherence of expressivity across language groups.

An interesting observation is L1 speakers for negative expressions (i.e., imposition (contempt, obviousness, irony) follow a strategy compatible with the Frequency code (and L1 listeners are expecting changes accordingly); on the contrary, L2 speakers appear to use the Effort code, which contributes to confusions in the audio-only modality. It appears that perceptual mismatches are linked to the expressive choices of speakers, which are in turn linked to their cultural origin.

References

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. Language and Speech, 3(4): 223–234.

Fujisaki, H. & Hirose, K. (1993). Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. ESCA Workshop on Prosody, 254-257.

Gussenhoven, C. (2004). The phonology of tone and intonation. Cambridge University Press.

Martins-Baltar M. (1977). De l'énoncé à l'énonciation: une approche des fonctions intonatives. Paris: Didier.

de Moraes, J. A. (2008). The pitch accents in Brazilian Portuguese: Analysis by synthesis. Speech Prosody, 389–397.

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In Hinton, L., Nichols, J. & Ohala, J. J. (Eds.), Sound symbolism, Cambridge: Cambridge University Press, 325–347.

Shochi, T., Rilliard, A., Aubergé, V. & Erickson, D. (2009). Intercultural perception of English, French and Japanese social affective prosody, In S. Hancil (ed.), The role of prosody in affective speech, Linguistic Insights 97, Bern: Peter Lang, AG, Bern, 31-59.

Wichmann, A. (2002). Attitudinal intonation and the inferential process. Speech Prosody, 11–16.

Audience design effects on prosody

Kathryn Weatherford & Jennifer E. Arnold (UNC Chapel Hill)

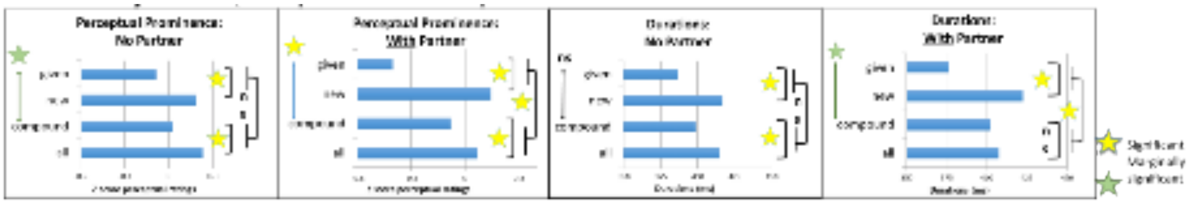
Information status is marked prosodically. For example, previously-mentioned "given" words tend to be acoustically reduced, while new information is not in focus and requires acoustic prominence [1,2]. Listeners are sensitive to this, shifting attention to new information in response to accenting [3]. This suggests that prosody is a selectional process used to communicate discourse status between conversational partners. In contrast, fluency accounts suggest that reduction is driven by speaker-internal mechanisms: words are reduced when they are easy to retrieve and articulate, regardless of discourse status information for the listener [4,5].

Our study examines whether the presence of a live, interactive addressee modulates variation in acoustic prominence. In particular, it tests the hypothesis that the relative contributions of selectional vs. facilitative mechanisms may be modulated by the social goals of the task. Given the importance of communicating information status [6], the presence of a communicative goal may encourage speakers to use greater variations in prosody to communicate these changes when a listener is present.

Following earlier work in our lab, we used a paradigm in which speakers described images of animals performing pairs of actions (spin, expand, blink, shrink), in four lexical conditions, with or without a conversational partner. The target was the second sentence:

1) Given: The panda spins. The **panda** blinks.
2) New: The frog spins. The **panda** blinks.
3) Compound: The panda and the frog spin. The **panda** blinks.
4) All: All the animals spin. The **panda** blinks.

Both selectional and facilitative accounts predict that "panda" will be reduced in the Given vs. New conditions. The critical condition is Compound: "panda" is out of focus (predicting prominence by the selectional account) but also facilitated due to previous articulation (predicting reduction by the fluency account). The All condition provides the comparison condition for the Compound condition, controlling for the multiple-animal context. If a conversational partner increases the need for successful communication of information status, then the acoustic prominence of the not-focused target words, particularly in the Compound condition, should be greater in the partnered vs. un-partnered condition. If acoustic prominence is internally driven, the presence of a partner should not matter.



Results (48 subjs) show that partners matter. We analyzed two metrics of acoustic prominence: 1) RA ratings (scale of 1-3), averaged over four RAS, and 2) log duration. In the No-Partner condition, both measures revealed that the Compound condition (compared to All) was just as reduced as the Given condition (compared to New). This supports the fluency account. In the presence of a partner, however, the Compound condition (compared to All) was significantly less reduced than the Given condition (compared to New). These results provide support for the simultaneous effects of both selectional and facilitation mechanisms. They also demonstrate that the presence of a live, interactive addressee can modulate variation in acoustic prominence, perhaps by shifting the relative contributions of selectional and facilitation mechanisms on prosody.

References

1. Breen, M., Fedorenko, E., Wagner, M. & Gibson, E. (2010). Acoustic correlates of information

structure. *Language and Cognitive Processes, 25(*7-9), 1044-1098.

2. 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*(5), 489-504.
3. Arnold, J.E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*(4), 495-527.
4. Bard, E. G., & Aylett, M. P. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language as-product and language-as-action traditions* (pp. 173–191). Cambridge, MA: MIT Press.
5. 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.
6. Ito, K. & Speer, S.R. (2011) Semantically-independent but contextually-dependent interpretation of contrastive accent. In S. Frota et al. (eds.), *Prosodic categories: production, perception and comprehension, studies in natural language and linguistic theory* 82 (pp. 69-92) Netherlands: Springer Science Business Media B.V.

****

Variability as a key feature of Autism spectrum disorders prosody

Ayako Kondo[1], Takayuki Konishi[2], Ken'ya Nishikawa[3], Hidetoshi Takahashi[4], Yoko Kamio[4], & Reiko Mazuka[3]
[1] The United Graduate School of Education, Tokyo Gakugei University, [2] Waseda University, [3] RIKEN Brain Science Institute, [4] National Institute of Mental Health, National Center of Neurology and Psychiatry

It has long been known that the speech of people with autism spectrum disorders (ASD) has atypical prosodic features. Yet how their prosody differs from that of typically developing (TD) persons has not been well-understood. In the present paper, we will show that the use of an intonational phonological framework provides tools to capture the nature of ASD prosody in a phonologically interpretable way and shed light on our understanding of ASD prosody.

Previously, we have found that ASD participants can produce lexically and syntactically determined aspects of prosody intact. In the present study, therefore, spontaneous speech of ASD and TD children are analyzed. Twelve children (age 7-17) diagnosed with high functioning ASD, and 14 TD children (age 7-16) were recorded while each child talked with an experimenter about common topics: e.g., hobbies and school life, for 5-10 minutes. Three trained phoneticians listened to the speech, and marked the sections that were judged "atypical." The atypical sections were then annotated using the X-JToBI scheme.

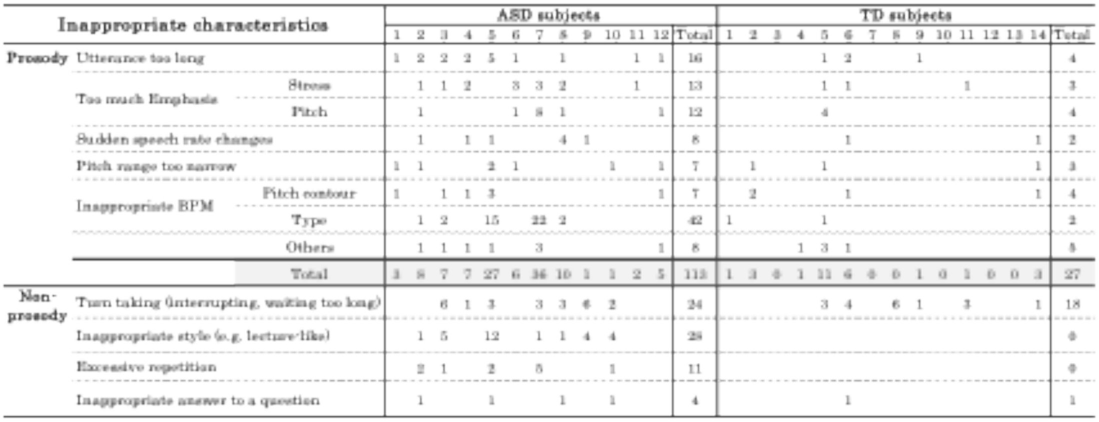
When the phoneticians listened to the entire recordings of individual children, the prosody of almost all of the ASD children was judged to "sound atypical," while none of TD children's was. Yet, the portion of the speech that was marked atypical constituted less than 10% of their speech. The fact that they were able to produce appropriate prosody 90% of the time indicates that they do not lack basic competence in producing various components of prosody.

To further analyze what kind of atypicality ASD children's speech contained, atypical sections of speech were classified into subtypes of prosodic and non-prosodic atypicality, as shown in Table 1. The results revealed two important characteristics.

First, although ASD children produced many more atypical prosodic and non-prosodic features than TD children, many of them appeared to be associated with the core symptoms of ASD in social, communicative interaction difficulty, e.g., problems with turn-taking and inappropriate use of BPM (tones that can occur at the end of an accent-phrase and contribute to the pragmatic interpretation of the phrase). Excessive repetition may be linked to another symptom of ASD: perseverance. It indicates that many aspects of ASD children's atypical speech may be the byproduct of their core ASD symptoms rather than deficits in prosody per se. At the same time, ASD children's use of amplitude for emphasis, even though pitch is used for emphasis in Japanese, is difficult to attribute to ASD symptoms. It may suggest that a problem in prosody itself may also exist in ASD.

Second, individual ASD children differed widely in their atypicality. Each ASD child showed atypical prosody in some ways, but where that atypicality manifested itself differed from child to child. Considering that the individual symptoms of ASD are highly variable, it is not surprising that ASD prosody is also highly variable. But it means that a conventional approach to characterizing ASD prosody by averaging across ASD individuals would not have given meaningful results. Instead, ASD prosody may be better understood by considering variability as a key feature.

Table 1 Classification of atypical portion of speech for ASD and TD children.



Inappropriate characteristics			ASD subjects													TD subjects														
			1	2	3	4	5	6	7	8	9	10	11	12	Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Prosody	Utterance too long		1	2	2	2	5	1		1			1	1	16					1	2			1						4
	Too much Emphasis	Stress		1	1	2		3	3	2			1		13					1	1					1				3
		Pitch		1				1	8	1				1	12					4										4
	Sudden speech rate changes			1		1	1			4	1				8						1								1	2
	Pitch range too narrow		1	1			2	1				1		1	7		1			1									1	3
	Inappropriate BPM	Pitch contour	1		1	1	3							1	7		2				1								1	4
		Type		1	2		15		22	2					42	1				1										2
		Others		1	1	1	1		3					1	8				1	3	1									5
		Total	3	8	7	7	27	6	36	10	1	1	2	5	113	1	3	0	1	11	6	0	0	1	0	1	0	0	3	27
Non-prosody	Turn taking (interrupting, waiting too long)				6	1	3		3	3	6	2			24					3	4		6	1		3			1	18
	Inappropriate style (e.g. lecture-like)			1	5		12		1	1	4	4			28															0
	Excessive repetition			2	1		2		5			1			11															0
	Inappropriate answer to a question			1			1			1		1			4						1									1

****

Native language effects on phrasal intonation: Application of smoothing spline analysis of variance (SS ANOVA) to multi-syllabic utterances

Tuuli H. Morrill (George Mason University)

When listening to foreign languages, people hear prosodic cues as they do in their native language, associating certain pitch and duration patterns with stressed syllables and prosodic boundaries (e.g., Cutler, Mehler, Norris, & Segui, 1992). But does native language prosody affect the production of non-native stress patterns and intonation contours with this kind of regularity? The current study demonstrates a novel application of a statistical technique, smoothing spline analysis of variance, to intonation contours produced by native and non-native speakers of English.

These data are from 197 speakers reading an identical speech passage in English (George Mason University's Speech Accent Archive (http://accent.gmu.edu)). The speakers' native languages are Mandarin (*n* = 50), Korean (*n* = 44), Arabic (*n* = 48), and English (*n* = 55), which differ in their prosodic structures. English intonation contours are typically shaped by tonal targets on accented syllables and at phrase boundaries, as well as the interpolated pitch changes between them (Beckman & Pierrehumbert, 1986). Thus, intonation contours produced by non-native speakers may be shaped differently if (1) accented syllables occur at different times for native and non-native speakers, and (2) high and low boundary tones are used differently by native and non-native speakers.

Smoothing spline analysis of variance (SS ANOVA) is used for comparing curves along multiple reference points, and researchers have started using SS ANOVA to examine pitch contours of syllables and words in tone languages (e.g., Yiu, 2014). Here, SS ANOVA is used to examine phrasal intonation contours. First, each speaker's recording was divided into phrases, and each phrase was divided into 1000 equally spaced time points at which an F0 value was extracted with Praat (Boersma & Weenink, 2012). F0 values were transformed to semitones relative to 1Hz and normalized by speaker. SS ANOVA was implemented with the gss package in R (Gu, 2014) and pitch contours were modeled with 95% Bayesian confidence intervals.

Analysis of these contours revealed consistent patterns based on native language group. Interestingly, native and non-native speakers placed higher-level prosodic boundaries at similar time points. However, language groups differed in the placement and degree of pitch extrema, as well as the realization of certain stylized contours. For example, Korean speakers showed transfer from native-language phrasal intonation patterns. Arabic speakers showed frequent F0 rises and falls consistent with repeated pitch accent patterning in some dialects of Arabic. Though Mandarin speakers produced tonal groupings and boundary tones similarly to native English speakers, they exhibited fewer clear F0 peaks corresponding to pitch accents. These results support a recent proposal by Jun (2014) for the inclusion of "macro-rhythm" in prosodic typology. Languages like Korean and Arabic exhibit strong macro-rhythm, with repeated tonal groupings at the prosodic word and intermediate phrasal levels, while languages like Mandarin exhibit weak macro-rhythm. The current results suggest that macro-rhythm could influence non-native speech production – in a medium-strength macro-rhythm language like English, non-native speakers from both ends of the scale differ from native speakers in ways consistent with their own native language.

References

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [Computer program]. (Version 4.0.26). Software and manual available online at http://www.praat.org.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

Gu, C. (2014). Smoothing Spline ANOVA Models: R Package gss. *Journal of Statistical Software*, *58*(5).

Jun, S.-A. (2014). Prosodic typology: by prominence type, word prosody, and macro- rhythm, in *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford University Press.

Yiu, S. (2014). Tone spans of Cantonese English. In *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014)* (pp. 143–146). Nijmegen, the Netherlands: ISCA Archive.

****

Lexical stress and the scope of boundary lengthening

Argyro Katsika (Haskins Laboratories)

It is well established that units of speech at the vicinity of prosodic boundaries are longer than their phrase-medial counterparts, a phenomenon known as boundary lengthening. However, current research has not clarified yet what factors determine the scope of this effect and how. One factor that gradually emerges as determinant is the position of the last lexical stress of the phrase, which seems to be relevant for the scope of boundary lengthening not only phrase-finally (e.g., [2], [4], [5]), but phrase-initially as well ([3]). However, the evidence is inconclusive, and even contradictory (compare [2] to [4]), not allowing solid conclusions on the exact role of lexical stress in determining the scope of boundary lengthening. This uncertainty might be partly due to a possible confound of the effect of lexical stress with the effect of pitch accent (but see [4]), and also partly due to the highly variable and speaker-specific manifestation of boundary lengthening (e.g., [1], [2]). To address these possibilities, an electromagnetic articulography (EMA) study of Greek was conducted that separately examined the lexical (lexical stress) and the phrasal (pitch accent) effects of prosody on boundary lengthening within a wide range of prosodic contours covering different boundary strengths and all the types of boundary tones used in Greek.

The results revealed a systematic effect of lexical stress phrase-finally regardless of whether the phrase-final word was accented or de-accented. Specifically, in stress-final words, pre-boundary lengthening scoped over the boundary-adjacent articulatory movements, meaning the movements of the consonant and vowel of the phrase-final syllable. In words with non-final stress on the other hand, lengthening was initiated further leftward from the boundary. Although the exact onset of the effect was speaker-specific, its dependence on the position of stress was robust, and remained stable across prosodic contours, boundary strengths, and boundary tones. This effect of stress was also congruous with the patterns shown by the pauses that followed the phrase-final words. These pauses involved specific articulatory configurations, where the spatially highest point (considered to be their articulatory target) was reached later in words with final stress than in words with non-final stress. However, the effect of the last lexical stress of the phrase did not spread across the boundary, i.e., it did not affect the initial part of the following phrase. To the contrary, phrase-initial boundary lengthening was consistently detected on the articulatory movements comprising the initial consonant, and was not affected by the lexical stress of the word preceding the boundary.

These results indicate that, although the scope of boundary lengthening presents high variability, which is largely speaker-dependent, the lexical aspect of prominence exerts a regular effect on it. Based on these results, a gestural account of prosodic boundaries is proposed in which lexical and phrasal prosody systematically interact. The implications of this account for prosodic structure and prosodic processing will be discussed. [Work Supported by NIH.]

References

[1] Byrd, D., Krivokapić, J. & Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America*, 123, 4456- 4465.
[2] Byrd, D. & Riggs, D. (2008). Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association*, 38, 187-202.
[3] Shattuck-Hufnagel, S. & Turk, A. (1998). The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America*, 103, 2889-2889.
[4] Turk, A. E. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35, 445-472.
[5] White, L. (2002). *English Speech Timing: A domain and locus approach*. PhD thesis The University of Edinburgh.

****

Dialect imitation across typologically distinct prosodic systems

Mariapaola D'Imperio[1] & James German[2]
[1] Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France & Institut Universitaire de France, Paris, France, [2] Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence

All phonological systems give rise to variability in the output, though this variability generally appears reduced when the right phonological generalizations are considered. Speakers attempting to imitate unfamiliar systems must learn which factors govern variability present in the target speech. This is expected to be more difficult when the native system and the target system are typologically distinct. This study explores variability in the intonational contours of Singapore English (SgE) speakers attempting to imitate American English (AmE), which has a typologically distinct prosodic system.

SgE and AmE largely overlap in terms of lexicon and grammar, and are generally mutually intelligible. They differ, however, in their intonational phonology. In SgE, the basic unit of intonation is the Accentual Phrase (AP), which is generally shorter than the AmE ip, and is characterized by L and H tones at its left and right edges (Chong 2012). Unlike AmE, there are no pitch accents, and tonal correlates of lexical stress only appear in utterance-final APs. Therefore, SgE more closely resembles edge-prominence languages like Korean (Jun 1993, 2005) than AmE. Given these strong typological differences, we ask how SgE speakers understand intonational variability present in AmE, and if and how they approximate alignment and scaling values of a model speaker.

We recorded 20 speakers of SgE, first in their native dialect, then while imitating an AmE speaker. SgE prosody differs among ethnic groups (Tan, 2010), so ethnically Chinese speakers were selected. The 36 target sentences included a three-syllable, initial-stress target word, which was either sentence-initial, at a continuation boundary, or sentence-final in a polar question. The AmE speaker produced the sentence-initial targets with a L+H* pitch accent on the initial syllable, followed by a fall to the L-boundary word-finally (Fig 1). The SgE pattern for the same target involves a rise from a L target at the beginning to a H peak near the end of the word (Fig 2). Approximating the AmE pattern therefore required adjusting the alignment of f0 peaks to a much earlier position. Since SgE does not have pitch accents (or other phrase-internal tonal landmarks), this task is predicted to be more difficult than when the two varieties are typologically similar (e.g., Australian English and AmE), where approximating the new system is primarily a matter of learning a new value for the proportional alignment of the f0 peak (i.e., a different implementation rule, D'Imperio et al. 2014). The continuation and question contexts similarly required adjustments to the timing of the L elbow. Additionally, there is strong downstep between the first and second AP in SgE, but only slight declination in the model's speech. Hence scaling adjustments were also expected. Analysis of 10 participants revealed that all speakers successfully adjusted peak alignment on a by-token basis, suggesting a strong role for phonetic detail. Alignment and scaling was adjusted for other target positions, though with different patterns of variability across speakers. We report on the structure of this variability vis-à-vis individual tokens and language background, and discuss its implications for the representation of phonological generalizations in late learning.

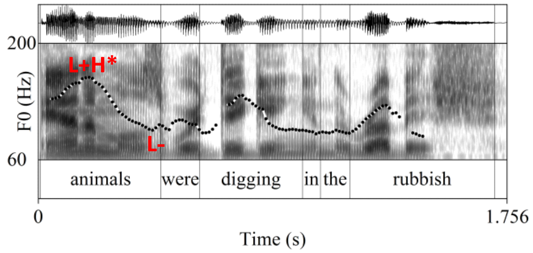


Figure 1. F0 contour for a target sentence as produced by the AmE model speaker.

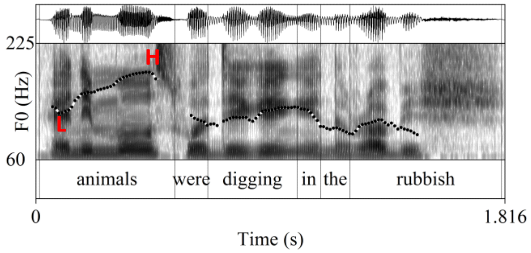


Figure 2. F0 contour as produced by one SgE speaker in the baseline task.

References

Chong, A.J. (2012). A preliminary model of Singaporean English intonational phonology. *UCLA Working Papers in Phonetics*, 111, 41-62.

D'Imperio, M., Cavone, R. & Petrone, C. (2014). "Phonetic and phonological imitation of intonation in two varieties of Italian". *Frontiers in Psychology*. 2014, 14 pages.

Jun, Sun-Ah (1993) *The Phonetics and Phonology of Korean Prosody*. Unpublished Ph.D. dissertation. The Ohio State University, Columbus, Ohio.

Jun, Sun-Ah (2005) Editor. Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press.

Tan, Ying Ying (2010). "Singing the same tune? Prosodic norming in bilingual Singaporeans." in M. Ferreira (ed), *Multilingual Norms*. Frankfurt: Peter Lang.