Perspectives on Rhythm and Timing

Glasgow, 19-21 July 2012

Book of Abstracts

English Language
University of Glasgow, Glasgow G12 8QQ, U.K.
e-mail: port.workshop@gmail.com
web: www.gla.ac.uk/rhythmsinscotland
Twitter: @RiS_2012 #PoRT2012

University of Glasgow

THE ROYAL SOCIETY OF EDINBURGH

The production and perception of English speech rhythm by L2 Saudi Speakers

Ghazi Algethami

University of York
ghazi.algethami@gmail.com

Keywords: L2 speech, rhythm, perception, metrics

1. INTRODUCTION

One of the major factors that contribute to the perception of foreign accent in the speech of L2 English speakers is the inappropriate use of rhythm [2]. However, it is one of the least studied aspects of L2 speech [1]. This is mainly because of the elusive nature of speech rhythm, and the lack of consensus among researchers about how to measure it.

Recently developed acoustic metrics of speech rhythm have offered a tool for L2 speech researchers to examine the production of speech rhythm by L2 speakers. The current study examined the production of English speech rhythm by L2 Saudi speakers by using a combination of two rhythm metrics %V and VarcoV (the percentage of vocalic segments and the standard deviation of vocalic segments divided by the mean and multiplied by 100, respectively) as they were shown to be complementary and successful in differentiating between languages rhythmically [3].

2. METHOD

Six intermediate and six advanced L2 Saudi speakers of English, as well as six native speakers of English, produced ten English utterances. Six native speakers of Saudi Arabic (a stress-timed language) also produced ten Saudi Arabic sentences for comparison. %V results showed that English was more stress-timed than both Saudi Arabic and L2 speaker groups, and there was no difference between Saudi Arabic and the L2 speaker groups. VarcoV results showed also that English was more stress-timed than Saudi Arabic; however, Saudi Arabic was more stress-timed than the L2 speaker groups. In both cases, there was no difference between the L2 speaker groups.

To ascertain the psychological aspects of these metric measurements, the four longest English utterances from the production task spoken by three native speakers of English, three advanced L2 speakers, and three intermediate L2 speakers were low-pass filtered at 300 Hz, and monotonized at 150 Hz, and then had their intensities neutralized at 75dB. These modified utterances were then presented to 23 native English listeners to rate them on a six-point scale according to the possibility of the utterance being produced by a native or non-native speaker.

3. RESULTS AND DISCUSSION

The listeners were able to differentiate between the native and nonnative speakers regardless of the highly distorted speech. There was a significant correlation between the perceptual ratings and the scores of VarcoV, but not %V.

4. REFERENCES

[1] Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

[2] Gut, U. (2003). Non-native speech rhythm in German. In *Proceedings of the15th International Congress of Phonetic Sciences* (pp. 2437-2439). Barcelona, Spain.

[3] White, L., & Mattys, S. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*, 501-522.

The role of rhythm classes in language discrimination

Amalia Arvaniti[a,b], *Tara Rodriquez*[a]

[a]University of California San Diego, [b]University of Kent
aarvaniti@ucsd.edu

Keywords: rhythm class, discrimination, pitch, timing, tempo

1. INTRODUCTION

The view that languages fall into distinct rhythm classes has been primarily supported by discrimination experiments like [1]. Discrimination, however, can be due to a number of factors. Here this possibility is explored by means of five AAX experiments that examine whether discrimination is due to rhythm class or to other prosodic properties, specifically tempo and F0.

2. METHOD

A different set of 22-24 listeners (all UCSD undergraduates) took part in each of the five experiments. The materials were sentences of English, Danish, Polish, Spanish, Korean and Greek, read by one male and one female speaker of each language (2 m and 2 f speakers for English which was the context and control). The sentences were converted into *sasasa* (i.e. all consonantal intervals were turned into [s] and all vocalic ones into [a]).

In each experiment, English was compared to one language under four conditions: the *sasasa* stimuli either retained the tempo (measured as vocalic intervals per second) of the original utterances (*T*) or were manipulated to have all the mean tempo of the stimuli of both languages in each experiment (*NoT*); F0 was either original (*F0*) or flattened (*NoF0*).

Table 1: rhythm classification and tempo differences between English (context) and targets; NDP stands for "Discrimination Not Possible", DP for "Discrimination Possible."

Language	Rhythm class	Tempo wrt English	Predictions Rhythm	Predictions Other
Danish	stress-timed	D ~ E	DNP	DNP
Polish	stress-timed	P > E	DNP	DP
Spanish	syllable-timed	S > E	DP	DP
Greek	syllable-timed	G > E	DP	DP
Korean	syllable-timed	K > E	DP	DNP

If discrimination is based on rhythm class, then it should accord with rhythm class differences (see Table 1). If it is due to tempo, discrimination should be possible whenever large tempo differences are present (see Table 1). Original F0 was hypothesized to aid discrimination in all experiments.

3. RESULTS

As Figure 1 shows, discrimination (A' > 50) was not based on rhythm class: Danish and Polish were discriminated from English but Spanish was not (results based on one-sample t-tests). Planned comparisons following an ANOVA on A' scores showed that discrimination depended on language and condition. For Danish discrimination was possible in T-F0 and noT-noF0. For Greek and Polish, discrimination was significantly better when original tempo was retained (T-F0, T-noF0), while for Korean, discrimination was worst when only F0 information was present in the stimuli (noT-F0).

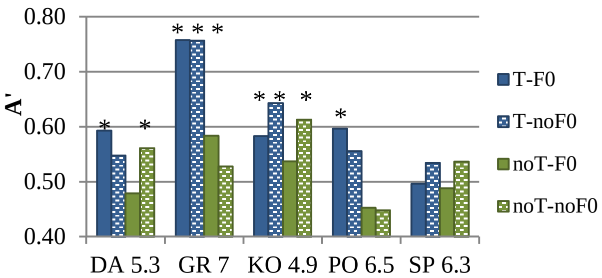


Figure 1: A' scores per language and condition; numbers next to language names refer to average tempo (in vocalic intervals per second). Stars indicate A' scores significantly different from chance.

4. DISCUSSION

These results strongly suggest that previous results interpreted as evidence for rhythm classes were most likely due to confounds between tempo and rhythm class. They further show that the *sasasa* transform is not ecologically valid: results differed depending on whether F0 was present, suggesting that the timing information encoded in *sasasa* is not processed independently of the other components of prosody. In conclusion, language discrimination in AAX experiments cannot be attributed to a single factor like rhythm but is most likely due to a combination of prosodic differences across languages.

5. REFERENCES

[1] Ramus, F., Dupoux E., Mehler. J. 2003. The psychological reality of rhythm class: Perceptual studies. *Proceedings of the 15th ICPhS*. 337-340.

A new look at subjective rhythmisation

Rasmus Bååth

Lund University Cognitive Science, Lund University, Sweden
rasmus.baath@lucs.lu.se

Keywords: meter; subjective rhythmisation

1. INTRODUCTION

Subjective rhythmisation (SR) is the phenomenon that when presented with a sequences of isochronous, identical sounds one can experience a pattern of accents that groups the sounds, often in groups of two, three or four [1]. SR has been explained using a model of rhythm perception based on neural oscillation [3]. The present study aims at extending the scope of earlier studies by including sequences from a wider range of inter stimuli intervals (ISI). A second aim is to investigate how perceived grouping relates to tempo, measured as the sequence inter stimulus interval (ISI). Using a simple model of neural oscillation as a starting point gives that the relation between grouping and ISI should follow a power-law distribution where grouping = c · ISI β , c and β being constants with β being negative.

2. METHOD

Nine female and 21 male participants, ranging in age from 19 to 78 years (M=31.6, SD=12.8) were recruited from the Lund community. The participants were seated in front of a computer wearing headphones. The task consisted of 32 click sequences of different tempo presented in a random order. The ISIs of the sequences were 150, 200, 300, 600, 900, 1200, 1500 and 2000 ms. For each sequence the participants were to indicate if they felt a grouping of the clicks on a scale ranging from “No grouping/groups of one” to “Groups of eight”. An online version of the the task can be found at http://www.sumsar.net/files/sr_task/public_sr_task.html.

3. RESULTS

All participants managed to carry out the task and for all participants but one, there was a significant negative correlation between ISI and perceived grouping. The participant that showed no significant correlation is not included in the subsequent analysis. Figure 1 shows the distribution of the probability of perceiving a grouping as a function of ISI. Groupings of five, six and seven are not included as they accounted for only 3% of the reported groupings. Figure 2 shows a log-log plot of the mean of perceived grouping as a function of ISI. The superimposed line shows the linear regression fit of the log-transformed data (β=-0.52, R^2=0.63, p<0.001).

Figure 1: The probability of perceiving a grouping by ISI.

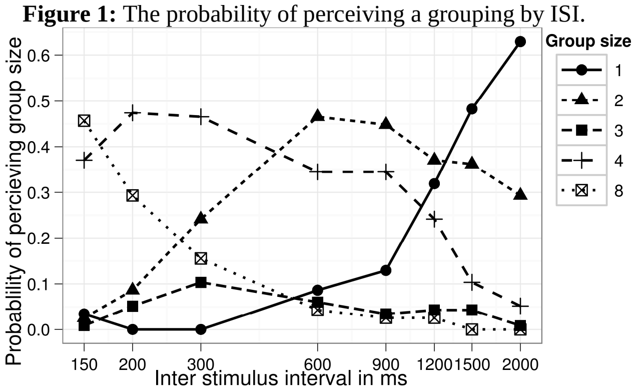
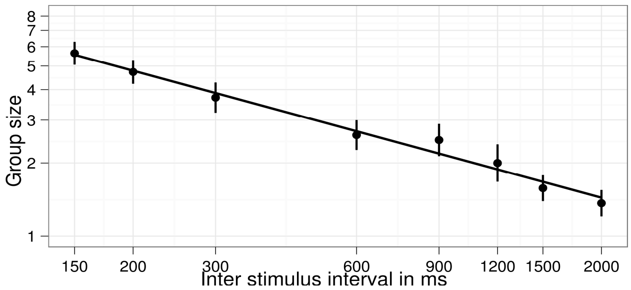


Figure 2: Mean grouping by ISI. The line ranges show the standard error.



4. DISCUSSION

This study replicates two of the findings of earlier SR studies, that is, perceived grouping tends to increase as ISI decreases and even groupings are much more common than odd. From ISI 900 ms to 2000 ms there is a steep increase in the probability to perceive no grouping, with a peak probability of 63% at ISI 2000 ms. This is in accordance with the notion of there being an upper limit to subjective rhythmisation but is discordant with an often proposed upper limit in the range of ISI 1500 ms to 2000 ms [2]. A power-law distribution appears linear on a log-log plot. Figure 2 shows an approximately linear relation and support the prediction, given by a simple model based on neural oscillation, that the relation between grouping and ISI follows a power-law distribution.

5. REFERENCES

[1] Vos, P. (1973). *Waarneming van metrische toonreeksen*. Nijmegen: Stichting Studentenpers.

[2] Fraisse, P. (1982). Rhythm and tempo. *The psychology of music*, 149–180.

[3] Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. *The psychology of time*, 189–231.

Native language and stimulus complexity affect rhythmic grouping of speech

Anjali Bhatara[1,2], Natalie Boll-Avetisyan[3], Annika Unger[3], Thierry Nazzi[1,2] and Barbara Höhle[3]

1. Université Paris Descartes, Sorbonne Paris Cité, France 2. CNRS (Laboratoire Psychologie de la Perception, UMR 8158), France 3. Universität Potsdam, Germany
bhatara@gmail.com

Keywords: rhythmic grouping, language experience, crosslinguistic comparisons, iambic/trochaic law

1. INTRODUCTION

Listeners group sequences of sounds in specific ways, depending on the acoustic variation in the sequence. Sequences of sounds that vary in intensity tend to be grouped in trochaic pairs, with the sound higher in intensity placed in the initial position, while sounds that vary in duration tend to be grouped in iambic pairs, with the longer sound placed in the final position [5]. These observations have been formulated into the Iambic-Trochaic Law (ITL) [3]. Pitch may also function in a manner similar to intensity, leading to trochaic grouping [4].

Effects of the ITL on rhythmic grouping preferences are assumed to be universal [3]. However, the world's languages differ as to their use of stress. Languages such as English and German utilize pitch, intensity and duration cues to communicate lexical stress, while French does not use lexical stress. Adult French speakers show a "deafness" to lexical stress [1]. This leads to the question: does linguistic experience with lexical stress affect rhythmic grouping? One study to explore this question comparing French and English speakers found no difference between the two groups [2]. We hypothesize that experience with lexical stress affects rhythmic grouping preferences, and that complex, speech-like stimuli are needed to show a cross-linguistic difference.

2. METHOD

We performed three experiments comparing French and German monolinguals. In Exp. 1, we presented participants with synthesized speech stimuli like those from the previous study [2] (the syllable /ga/ repeated, varying in either intensity or duration). We asked participants whether they perceived strong-weak or weak-strong pairs. In Exp. 2, we used more complex stimuli comprising varied CV syllables, again varying in intensity or duration. In Exp. 3, we tested participants on the same sequences as Exp. 2, but this time varying in pitch or duration.

3. RESULTS

In Exp. 1, we found results similar to those from the previous study [2] with no difference between linguistic groups. In Exp. 2, we showed that both French and Germans showed grouping according to the ITL, but there was an interaction between native language and type of acoustic manipulation, with the Germans responding trochaic more often for the intensity-varied sequences and iambic more often for the duration-varied sequences. In Exp. 3, we found similar results for duration, but the French participants did not consistently group the pitch-varied sequences trochaically, though the Germans did.

4. DISCUSSION

These results demonstrate that native language has strong effects on rhythmic grouping according to the ITL, but it may be necessary that the stimuli are sufficiently complex to bring out differences between linguistic groups. In addition, pitch appears to be a more ambiguous cue for grouping than intensity, being perhaps more variable across languages or more dependent on language experience.

5. REFERENCES

[1] Dupoux, E., Peperkamp, S., Sebastian-Galles, N. 2001. A robust method to study stress “deafness.” *JASA*, 110, 1606–1618.

[2] Hay, J. S. F., Diehl, R. L. 2007. Perception of rhythmic grouping: testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122.

[3] Hayes, B. 1995. *Metrical stress theory: principles and case studies*. University of Chicago Press.

[4] Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., Donati, C. 2008. Different phrasal prominence realization in VO and OV languages. *Lingue e linguaggio, 7*(2), 1–28.

[5] Woodrow, H. 1909. A quantitative study of rhythm: The effect of variations in intensity, rate, and duration. *Archives of Psychology*, 14, 1–66.

Rhythmic regularity revisited: Is beat induction indeed pre-attentive?

Fleur Bouwer (1,2), Titia van Zuijen (1), Gábor Háden (1,2), Henkjan Honing (1,2)

1) Cognitive Science Center Amsterdam, University of Amsterdam, The Netherlands
2) Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands
f.bouwer@uva.nl

Keywords: rhythm, meter, attention, expertise, event-related potentials

1. INTRODUCTION

Beat is a psychological construct that allows us to hear regularity in music in the form of metrical structure. As such, beat perception is considered fundamental to music processing [1]. Regular external events can give rise to the feeling of a beat, through the process of *beat induction*. Beat induction is guided by accents in various forms and thus, is driven by sensory input. In addition to these external events, purely internal processes can influence how we perceive a beat.

Previous work has suggested that the brain can process the beat without attention, by showing that irregular events in strong metrical positions are processed different from irregular events in weak metrical positions [2]. However, in this study events in strong metrical positions were also acoustically different from events in weak positions. Therefore, the results might have been due to acoustic differences rather than differences in metrical position. Also, results from a recent fMRI study suggest that attention is a prerequisite for beat perception [3]. The current study therefore examines whether beat induction should indeed be considered pre-attentive, and investigates in how far not only structural but also acoustical factors influence the response to a metrical rhythmic sequence.

2. METHOD

EEG is commonly used to measure pre-attentive brain responses. We examine both event-related brain responses and oscillatory activity. We present subjects with a rhythmic sequence in which deviants occur in the form of infrequent sound omissions. We compare the ERP in response to an omission in a metrically strong position with the response to an omission in a metrically weak position, while controlling strictly for acoustic differences. Furthermore, we examine whether oscillatory activity in the beta and gamma range could function as an alternative index of pre-attentive rhythmic expectancy. The experiment is conducted with musicians and non-musicians to study the effects of expertise.

3. RESULTS

Preliminary results show that the amplitude of the MMN response to deviants in metrically strong positions is bigger than the amplitude of the MMN to deviants in metrically weak positions. This suggests that indeed, beat induction is pre-attentive. However, an exploratory analysis of the results shows that we cannot completely rule out an explanation of the results in terms of the acoustical structure of the stimuli.

4. DISCUSSION

To incorporate these and previous results, we propose a framework to study the relationship between beat perception and selective attention (Bouwer & Honing, in prep.). We introduce two hypotheses about beat perception and attention. The first hypothesis states that there can be no beat perception without attention. The second hypothesis entails that while beat induction is independent of attention, the perception of a beat can be modulated by attention through the influence of attention on the internal processes involved in beat perception.

5. REFERENCES

[1] Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, 1252: The Neurosciences and Music IV — Learning and Memory, 85–91.

[2] Ladinig, O., Honing, H., Háden, G. P., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception*, *26*, 377-386.

[3] Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J. A., Steinberg, F., & Large, E. W. (2010). Neural responses to complex auditory rhythms: The role of attending. *Frontiers in Psychology, 1,* 1-18.

Does musical training influence how we process spoken and hummed speech? An electrophysiological study using the Closure Positive Shift

Paulo Branco, Susana Silva, Fernando Barbosa, João Marques-Teixeira and São Luís Castro

Faculty of Psychology and Education, University of Porto, Portugal
slcastro@fpce.up.pt

Keywords: Closure Positive Shift; ERPs; musical expertise

1. INTRODUCTION

The Closure Positive Shift, CPS, is a component in event-related potentials that reflects the segmentation of sentences across prosodic boundaries [1, 2]. Its topography and latency vary according to stimulus modality (auditory or visual), lexical complexity (hummed or spoken sentences), and even listener age. A similar component is elicited when musicians listen to music containing boundaries between musical phrases [3]; this has been termed the music CPS, that differs somewhat from the language CPS in latency and topography. Our question here is whether musical expertise affects the CPS in speech. We compare musicians and non-musicians in the segmentation of spoken and hummed sentences using ERP methods and a probe detection paradigm.

2. METHOD

Participants were 16 musicians (at least 7 years of musical training; 9 women, mean age 20.2 yrs SD 2.5) and 16 non-musicians (10 women, mean age 20 yrs SD 3.5), all right-handed. They listened to short sentences and were asked to identify whether a probe word was presented. The sentences contained one or two phrase boundaries, and were spoken or hummed (2 x 2 x 45). For the probe detection task in the hummed condition, a spoken word was spliced onto 30 additional hummed sentences that served as fillers. EEG was recorded continuously from 200 ms pre-stimulus.

3. RESULTS

The ERP traces differed in sentences with one versus two boundaries, with evidence of CPS over one boundary contrasting with no CPS in the corresponding portion of sentence without the boundary ($p < 0.05$). Hummed sentences had a more frontal distribution than spoken ones. There was no effect of musical expertise in the amplitude of the CPS. In spoken sentences, the musicians had shorter onset latencies than non-musicians ($p < 0.01$; 0-300 ms window). In hummed sentences, non-musicians had a negative deflection in the 100-300 ms window, musicians did not ($p < 0.01$).

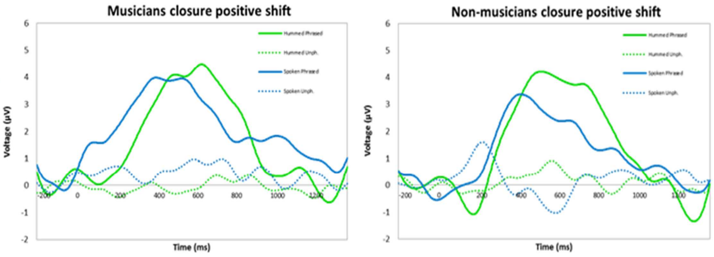


Figure 1: Closure positive shift, CPS, in musicians (left) and non-musicians (right). Spoken (blue) CPS has earlier latency in musicians. The negative deflection preceding hummed (green) CPS occurs only in non-musicians.

4. DISCUSSION

The early negativity found for hummed speech in non-musicians is consistent with Pannekamp et al's findings [2]. It probably signals the cognitive preparation required to process delexicalized speech. Our finding that musicians did not show this early negativity suggests that the absence of lexical information does not elicit such cognitive preparation, and thus that processing hummed speech is less demanding for them than for non-musicians. When lexical cues are present (spoken sentences), musicians seem to process prosodic boundaries faster than non-musicians. Taken together, these findings indicate that musical expertise may facilitate phrasing processes in speech.

5. REFERENCES

[1] Steinhauer K, Alter K, Friederici AD. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191–196.

[2] Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD. (2005). Prosody driven sentence processing. *Journal of Cognitive Neuroscience, 17*, 407–421.

[3] Knösche TR, Neuhaus C, Haueisen J, Alter K, Maess B, Witte OW, Friederici AD. (2005). The perception of phrase structure in music. *Human Brain Mapping, 24,* 259–273.

Metrical ambiguity in a piece of rhythmic music

Daniel Cameron[1], Job Lindsen[2], Marcus Pearce[3], Geraint Wiggins[3], Keith Potter[4] and Joydeep Bhattacharya[2]

[1]Brain and Mind Inst., University of Western Ontario, Canada [2]Dept. of Psychology, Goldsmiths, University of London, UK [3]Centre for Digital Music, Queen Mary, University of London, UK [4]Dept. of Music, Goldsmiths, University of London, UK
danielcameron01@gmail.com

Keywords: rhythm processing; EEG; inter-trial phase coherence; metric ambiguity

1. INTRODUCTION

Metre is the hierarchical organisation of rhythm, corresponding to dynamic perceived strength of positions in rhythmic sequences, and tactus, a frequency conventionally perceived as the 'beat' to which listeners readily synchronise movements. Metre perception and the endogenous process by which it emerges from rhythm are essential cognitive aspects of music, yet its underlying dynamical neural responses remain largely uncharacterized. While previous research has investigated event-related EEG aspects of metre perception [1,3], we explore spectral power (characterizing local neuronal synchronization), and inter-trial coherence (ITC) (characterizing the stimulus-locked, trial-to-trial phase consistency of brain responses) as they vary with musical rhythms which differ in metric information. We investigate a range of EEG frequency bands with special emphasis on low frequencies corresponding to the tactus of plausible metres of stimulus rhythms.

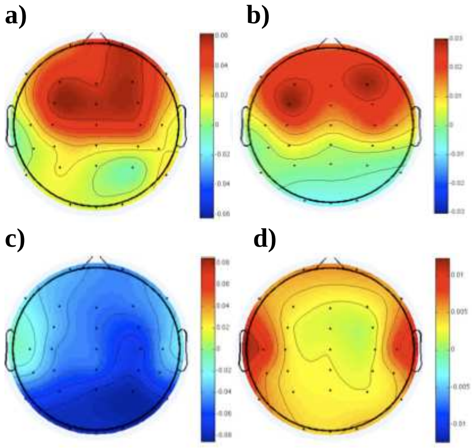
2. METHOD

Multivariate (32 channels) EEG signals were recorded from twenty musicians listening to two versions of Steve Reich's *Clapping Music* (1972), a piece consisting of two performers clapping a rhythm and systematically transforming it to produce 12 unique rhythms, each repeated 12 times. Rhythms are categorized according to a model of metre-induction [4] as having either an unambiguous metre of 3 tactus beats (1.33 Hz) or ambiguously plausible metres of 3 or 4 tactus beats (1.77 Hz for the latter), respectively. These different rhythm categories are compared in terms of their mean spectral power and ITC values of standard EEG frequency bands and also for specific low frequencies.

3. RESULTS

Metrically ambiguous rhythms, as compared to unambiguous rhythms, are associated with greater power at 1.33 Hz and 1.77 Hz, and in the broader delta band (1-4 Hz). These low frequency effects are most conspicuous over bilateral frontal areas (see Fig. 1a). Ambiguous figures also show greater power in the gamma band (24-60 Hz) over bilateral temporal regions (see Fig. 1d). On the other hand, unambiguous rhythms show greater power in the alpha band (8-12 Hz) (see Fig. 1c). Ambiguous figures are also associated with greater ITC at 1.33 Hz, 1.77 Hz, and the broader delta band, also over bilateral frontal areas (see Fig. 1b).

Figure 1: Topographical scalp distribution of the differences between metrically ambiguous and unambiguous rhythmic figures in a) delta power, b) delta inter trial phase-coherence, c) alpha power, and d) gamma power.



4. DISCUSSION

Differences in EEG power and phase (ITC) at the frequencies corresponding to plausible metre/tactus frequencies may reflect aspects of the endogenous perception/processes of metre. Localisation over premotor areas is consistent with past research on the role of motor systems in rhythm/metre perception and cognition [2]. Gamma band power differences over auditory cortex may correspond to metre-based differences found previously [1]. Rhythm structure may play a role, as effects correlate linearly with number of rests in rhythms. (i.e. rests may provide metrical information in the context of this rhythmic music, influencing EEG correlates of metre perception/processing).

5. REFERENCES

[1] Fujioka, T., Trainor, L.J., Large, E.W., & Ross, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of the New York Academy of Sciences*, 1169: 89-92.

[2] Grahn, J.A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, 19(5): 893-906.

[3] Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, 31(28): 110234-10240.

[4] Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4): 411-440.

Rhythmic priming of speech for enhanced processing

Nia Cason and Daniele Schön

INSERM U 1106, Institut de neurosciences des systèmes, Marseille, France
nia.cason@etumel.univmed.fr

Keywords: speech, music, rhythmic priming, perception, EEG

1. INTRODUCTION

A regular auditory rhythm induces rhythmic entrainment. This allows temporal expectations to be formed, and for attentional resources to become allocated in time [5]. The processing of both non-speech [4] and speech [7] is heightened at attended timepoints. These findings are relevant when considering how rhythmic stimuli may be used to enhance speech processing.

We hypothesised that using a rhythmical prime would enhance speech perception when:
1) **Beat Expectations** are met
2) **Metrical Expectations** are met

2. METHOD

- Behavioural and EEG experiment.
- 20 native French speakers, unimpaired speech.
- Stimuli: A Rhythmic Prime (Binary/Ternary metre) was followed by a pseudoword (bisyllabic/ trisyllabic) which contained a target phoneme.
- Word processing was measured by reaction times to a phoneme detection task.
- 2 x 2 Experimental Design: Beat conditions ('ON-beat' or 'OFF-beat' target phoneme) x Metric conditions ('Matched' or 'Mismatched' number of syllables to the Prime beats per IOI).

3. RESULTS AND DISCUSSION

- Behavioural Results: Reaction Times

'On-beat' targets < 'Off-beat' targets ($p<0.001$).

- Electrophysiological Results:

Metrical Expectations: A larger N100 amplitude in the presence of a Metrical violation (Figure 1a). This was much like the Mismatch Negativity response, elicited on detecting a 'mismatching' number of syllables (to the Prime's metric structure) [3] [6]. As this was an early component, it may be related to sensory processes.

Beat Expectations: There was a larger P300 amplitude for 'Beat' violations (Off-beat trials) (Figure 1b). This could reflect the detection of a low-probability (Off-beat) event [2] and it may index an impact on cognitive processes. A similar effect has been found in visual temporal orienting [1].

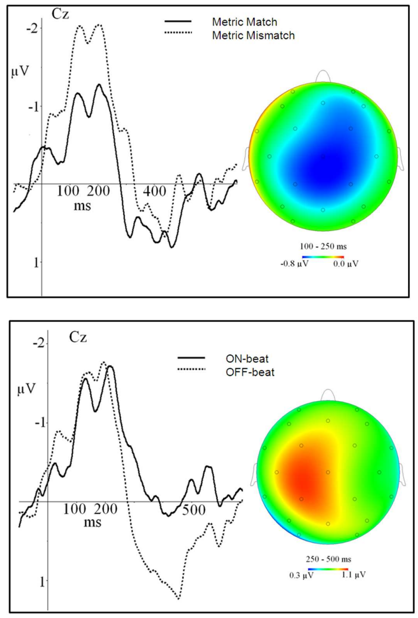


Figure 1: Top: a larger N100 amplitude in response to a metric violation and difference waves map. Bottom: a larger P300 amplitude in response to a beat violation and difference waves map.

- Broad Conclusion: Rhythmical stimuli which induce expectations about Beat and Metre might be used to enhance speech perception/production in clinical populations with hearing/speech impairments.

4. REFERENCES

[1] Correa A., Lupiañez J., Madrid E, Tudela P. (2006). Temporal attention enhances early visual processing: a review. *Brain Res* 1076:116-128.

[2] Donchin, E., Karis, D., Bashore, T., Coles, M., & Gratton, G. (1986). Cognitive psychophysiology: systems, processes, and applications. In: Psychophysiology: systems, processes, and applications. *New York: The Guilford Press* 244-267.

[3] Herholz, S. C., Claudia Lappe, C., Knief, A. & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise; *European Journal of Neuroscience* 28(11), 2352–2360.

[4] Jones, M. R., Moynihan, H., MacKenzie, N. & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* 13, 313–319

[5] Large, E. & Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review* 106(1), 119–159.

[6] Näätänen R., Paavilainen P., Alho K., Reinikainen K., Sams M. Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neurosci Lett.* 27:217–221.

[7] Quené, H. & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*, 1-13.

Tempo and musical performance: A tensive hearing

Marina Maluli César

Faculty of Letters, Philosophy and Human Sciences of the University of São Paulo, Brazil
marina.maluli@gmail.com

Keywords: musical semiotics, rhythm, tempo perception, musical performance

1. INTRODUCTION

The Greimas' semiotics line developed itself around the mechanisms responsible for definitions of action, focusing on storytelling as well as the analysis of discrete and binary contents. Little by little, we can notice a movement based upon the theory of modalizations from the act of doing to the being, while prioritizing the affective universe [1]. This new approach allowed the development of several semiotic's theories, among then we focus on the tensive semiotic found in Jacques Fontanille and Claude Zilberberg for instance [2]. To the tensive semiotic the sensible contents come to light, with allow us to study discoursive phenomena related to continuity. It was possible for instance to determine the fundamental tensive-phoriques flows of the generative trail. From this point on we can say that there is some kind of "music in semiotics" in the sense that concepts like rhythm, valence, vector and perception play an important role [3].

2. METHOD

The objective of our research is to study the changes of meaning occurred in a given composition, taking into consideration its interpretation during a musical performance [3]. To be more precise, we will analyze the changes of meaning generated by variations in tempo from different musical performances based on the composer's timing marks in the score. We believe that by using a tensive hearing analysis, i.e. an analysis based upon the Tensive Theory, it is possible to apply its theorical tools to musical discourses, which are born from the interaction between sound and time [6]. Besides that, we will discuss the influence that the choice of a specific musical instrument, as well as the role of gestures in musical performance, have on the connotative meaning of the composition [2].

3. REFERENCES

[1] Greimas, A. J. Pour une théorie des modalités. Langages 43, 1976.
[2] Hjelmslev, L. Prolegomena to a theory of language. University of Wisconsin Press, 1961.
[3] Novaes, G. Guiomar Novaes Plays Chopin. VOX, 1995. CD 1
[4] Zilberberg, C. Essai sur les modalités tensives. Amsterdam: Benjamins,1982.
[5] Zilberberg, C.Précis de grammaire tensive. Tangence, n° 70, automne 2002.
[6] Wisnik, J. M. O som e o sentido. São Paulo: Companhia das Letras. 1989.

Speech cycling in Korean

Younah Chung[a] *and Amalia Arvaniti*[a,b]

[a]University of California San Diego, [b]University of Kent
yachung@ucsd.edu

Keywords: rhythm, timing, cycling, Korean, rhythm class

1. INTRODUCTION

Korean has not been unanimously classified for rhythm class [1]. It also lacks stress [3] and thus it does not fit into views that speech rhythm rests on alternations of metrical strength. The aim here was to examine what, if any, elements are used in Korean for rhythm purposes. We used the speech cycling task which involves speakers repeating a phrase in time with a metronome [2]; under these conditions, speakers keep metrically prominent elements in stable phase [2]. It was hypothesized that in Korean the initial syllables of accentual phrases (APs [3]), act as beats playing the part stress plays in English. If so, the phase of syllables in cycling should be determined by their position in an AP (first or other), not their order in the utterance. But if Korean is syllable-timed [4], then cycling should reveal no other regularities beyond those due to syllable order.

2. METHOD

The materials were 12 sentences (mixed with 12 fillers). Each was 9 syllables and three APs long but the number of syllables in each AP varied (3-3-3, 2-4-3, 2-2-5); e.g. (i) [norɛga]$_{AP}$ [nɔmuna]$_{AP}$ [nɨɾida]$_{AP}$ 'the song is too slow' vs. (ii) [nabi]$_{AP}$ [nɛmariga]$_{AP}$ [naɾat'a]$_{AP}$ 'four butterflies flied'. Each AP pattern was tested with one phrase containing only CV and one containing mostly CVC syllables; compare (ii) above with (iii) [naldo]$_{AP}$ [nuŋnukʰago]$_{AP}$ [mudɔpt'a]$_{AP}$ 'The weather is also humid and hot.'

Five male and five female native speakers of Seoul Korean, in their 20s or early 30s repeated each sentence as many times as they could in one breath, fitting each repetition into beat intervals. The metronome was set at 32, 36 and 40 beats per minute. Each AP was expressed as a ratio of the entire cycle (d) as illustrated in Fig. 1: i.e. a/d is the phase of AP1, b/d that of AP2 and c/d that of AP3.

3. RESULTS

As can be seen in Fig. 2, although phase differences are present when the number of syllables in an AP differs, by and large phase is stable, especially across speaking rates. Syllable composition also affected phase, with APs with CVC syllables showing later phase (data not shown).

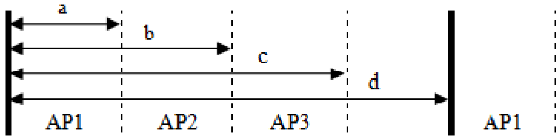


Figure 1: Measurements: a/b, a/d, b/d and c/d. Thick lines represent phrase onsets; broken lines AP boundaries.

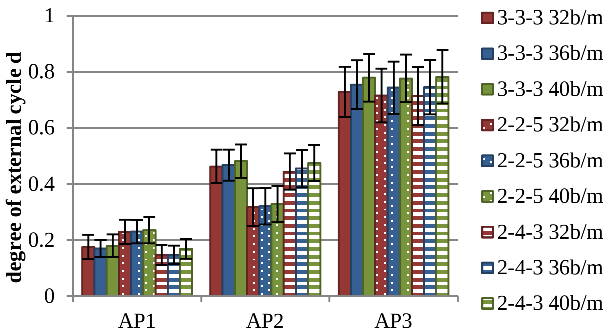


Figure 2: Mean phase and standard deviations of AP1, AP2 and AP3 separately for each metronome rate and AP composition (N = 10).

4. DISCUSSION

The results suggest that speakers keep the onsets of accentual phrases in phase although syllable count and composition also affect phasing. These results support our hypothesis that in Korean, AP onsets operate similarly to stresses. At the same time, the relative importance of the syllable cycle supports a view of rhythm that does not rest on the timing of one prosodic constituent, such as the AP or the stress foot, but on the relative salience of different levels of prosodic structure.

5. REFERENCES

[1] Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46-63.

[2] Cummins, F. & R. Port. 1998. Rhythmic constraints on stress timing in English. *JPhon* 26, 145-71.

[3] Jun, S-A. 2005. Korean intonational phonology and prosodic transcription. In S-A. Jun (Ed.), *Prosodic typology*, pp. 201-29. Oxford: Oxford University Press.

[4] Kim, J., Davis, C., & Cutler, A. 2008. Perceptual tests of rhythmic similarity: II. Syllable rhythm. *LangSp* 51, 343-59.

Does joint entrainment lead to social facilitation in infants?

Laura K. Cirelli, Kathleen M. Einarson and Laurel J. Trainor

Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada
cirelllk@mcmaster.ca

Keywords: entrainment, development, interpersonal synchrony, social behaviour

1. INTRODUCTION

Musical behaviour is unique to humans and cross-culturally universal. One fundamental music skill is the ability to entrain movement to an external auditory beat. It has recently been shown that the presentation of an isochronous auditory beat while participants do not move results in amplitude changes in oscillatory beta band (15-30Hz) activity following the beat in both auditory cortex and motor networks [1], suggesting a mechanism for auditory-motor entrainment.

Interpersonal entrainment is particularly important, as it has been associated with social bonding. For example, individuals who walk or sing together are subsequently more helpful and cooperative in later interactions with one another [2]. Evidence for this effect has also been shown in 4-year-olds [3].

The time course for the development of entrainment and its effects on social development are still unclear. When infants are bounced to music, it does affect their perception of the rhythm [4]. However, although infants tend to spontaneously produce rhythmic movement when listening to music [5], it tends to be asynchronous until preschool age.

The current study investigated whether the social effects of interpersonal entrainment are measurable in 14-month-old infants when the experimenter controls their movement.

2. METHOD

Each infant (tested individually) heard a melody containing either predictable (isochronous) or unpredictable (random) beats. Experimenter A (E-A) held the infant facing Experimenter B (E-B), and bounced the infant in a way congruent to what the infant was listening to (predictable beats or unpredictable beats). E-B bounced to beats that were either synchronous or asynchronous to the infant's movements. Thus, each infant was randomly assigned to one of the four conditions: predictable vs. unpredictable beats crossed with interpersonal synchrony vs. asynchrony. Infants were then tested in a situation in which they had the opportunity to help E-B by handing over desired objects following the methods of [6].

3. RESULTS

Infants who were bounced synchronously with the researcher while listening to predictable beats were significantly more helpful (61% helping likelihood) than those that were bounced asynchronously while listening to unpredictable beats (25% helping likelihood), $t_{(20.5)}=3.02$, $p=.007$. Infants in the other two conditions are currently being tested.

4. DISCUSSION

These results suggest that 14-month-olds do experience social facilitation following interpersonal auditory-motor entrainment. If this effect is a function of joint action alone and not dependent on auditory-motor entrainment, then we expect that the infants in the 'synchronous movements to unpredictable beats' condition will show helping rates similar to those in the 'synchronous movements to predictable beats' condition. If, however, it is interpersonal *entrainment* specifically that is required for these effects to occur then we expect that only the infants in the 'synchronous movements to unpredictable beats' condition will demonstrate increased rates of helping.

5. REFERENCES

[1] Fujioka, T., Trainor, L.J., Large, E.W., & Ross, B. in press. Internalized timing of isochronous sounds is represented in temporal dynamics of neuromagnetic beta oscillations in sensorimotor networks. *J. Neuroscience*.

[2] Wiltermuth, S.S., & Heath, C. 2009. Synchrony and Cooperation. *Psychological Science* 20, 1-5.

[3] Kirschner, S., & Tomasello, M. 2010. Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behaviour* 31, 354-364.

[4] Phillips-Silver, J., & Trainor, L. J. 2005. Feeling the beat in music: Movement influences rhythm perception in infants. *Science, 308*, 1430.

[5] Zentner, M., & Eerola, T. 2010. Rhythmic engagement with music in infancy. *PNAS,* 107, 5768-5773.

[6] Warneken, F., & Tomasello, M. 2007. Helping and cooperation at 14 months of age. *Infancy* 11, 271-294.

A gentle introduction to Dynamical Systems Theory for researchers in speech, language and music

Fred Cummins

University College Dublin
fred.cummins@ucd.ie

Keywords: Dynamical Systems Theory

1. INTRODUCTION

Dynamical Systems Theory (DST) is increasingly becoming the lingua franca of a diverse set of approaches that seek to analyze, model, and understand human behavior. Within these areas, the introductory materials introducing the concepts and techniques of DST found in textbooks for engineers are often of little help. In this tutorial, I will attempt to provide a basic foundation in DST that is of potential use for researchers in speech, language, and music. I will not assume any prior familiarity with DST, nor any mathematical expertise beyond basic school mathematics.

2. OVERVIEW

We start with the notion of *state*, whereby the modeller makes a choice of numerical variables that, it is hoped, provide an insightful index into the behaviour or temporal characteristics of some system of interest. It is important to recognize the responsibility of the modeller in choosing how state is to be indexed, and the implications of appropriate and inappropriate choices of state variables.

We are interested in the lawful, i.e. deterministic, evolution of state over time. This is expressed in a *dynamic*. An explicit expression of a dynamic will only be available for very simple systems, but these simple systems can help us to develop the theoretical arsenal and set of concepts that will be of use in describing complex systems too.

The qualitative features of a dynamical system will be of particular interest. These will include the number and type of long-term behaviors the system exhibits, its structural stability as a function of parameter change, and its stability in the face of external perturbation. We will distinguish between short-term *transient* behavior and long-term *attractors.*

The consequences of interaction among several dynamical systems are studied under the headings of *coupling* or *entrainment*. We will look at some of the better known forms of interaction, with a keen eye to the relevance of these concepts for those studying the rich but tricky notion of rhythm. Differences between rhythm in speech and language may be drawn out here. One or two variants on the theme of oscillators will be met, but without a great deal of detail.

Despite the limited time we have available to us, I hope we can touch on enough core concepts to whet the appetites of attendees.

3. REFERENCES

Any of the following might be found useful texts to complement this tutorial.

[1] Norton, A. (1995). Dynamics: an introduction. In Port, R. F. and van Gelder, T., editors, Mind as Motion: Explorations in the Dynamics of Cognition, chapter 1, pages 45–68. Bradford Books/MIT Press, Cambridge, MA.
(Introduces basic mathematical foundations within a volume that illustrates a wide variety of dynamical approaches to topics in cognitive science)

[2] Abraham, R. and Shaw, C. (1983). Dynamics, The Geometry of Behavior, Part 1. Aerial Press, Santa Cruz, CA.
(Graphical coverage of topics from basic to advanced, without any equations at all. Excellent for developing intuitions about the basic concepts of dynamic systems)

[3] Pikovsky, Arkady, Rosenblum, Michael and Kurths, Jürgen (2001). *Synchronization: A Universal Concept in Nonlinear Sciences,* CUP.
(More rigorous treatment of the mathematics, but very clear and methodical, with emphasis on concepts rather than proofs or techniques. Chapters 1—3 especially recommended)

[4] Kelso, J. A. Scott (1995). *Dynamic Patterns*, MIT Press.
(Kelso’s Coordination Dynamics is probably the best worked example of the rigorous application of dynamic systems to topics in cognitive science.)

Towards a musicology of temporal cycles

Kevin Dahan

Université Paris-Est Marne-La-Vallée – LISAA – EA 4120, France
kevin.dahan@{wanadoo.fr,univ-mlv.fr}

Keywords: musicology, music cognition, computer-assisted musicology, systematic musicology

1. INTRODUCTION

As music is an art of time, segmentation and subdivision play an essential part in it. There are different levels of understanding of these processes: many analytical methods deal with them by creating different levels of organization, from melodic elements to mesostructures and macrostructures [5]. In Schenkerian analysis, rhythm is considered as an element of the musical foreground, and therefore not considered as central in the determination of the *Ursatz*, the main structure to be determined by analysis [6]. Similarly, rhythm is absent from pitch-class sets based theories, only to be introduced at a later stage [2].

By defining rhythm as cycles of different lengths (and iterations), it is possible to consider musical activity as layers upon layers of different time cycles – of different rhythms and oscillations. To facilitate understanding, we will differentiate between the traditional rhythm (TR) conception and the layers of rhythmic cycles (LRC) approach. This new subdivision of the musical discourse allows us to draw parallels between different practices (for example, improvisation and composition), traditions (e.g. western, Indian traditional, jazz), as well as inside works themselves, as exemplified in [1]. In effect, the LRC approach integrates time with pitch, instead of considering it another dimension of musical expression that is articulated in a later stage, as is usual in the writings of composers and musicologist (see for example the inside/outside-time dichotomy in Xenakis' writings [7]).

2. AIMS

Developing an analytical model that is not limited to hierarchical classification of TR structures would allow for a more dynamical approach to musical understanding, allowing trans-level approaches to the musical phenomenon. From the frequency level of the pitch to "social" considerations in improvisation, the LRC approach is helpful in finding echoes where other methods are forced to rely on different representational strategies and models to be able to describe accurately the structural similarities and differences across levels.

The relations between music perception and cognitive aspects are already well-known from a neurological perspective [3]. What is lacking is a generalization of this concept with a musicological perspective.

3. TOOLS

This research should be based on three domains: musicology, cognition, and computer modeling. A thorough analysis of musical patterns (from melodic aspects *à la* Schenkerian, pitch-set based approaches, to a more complex examination of LRC variations in free improvisation setting, for example) is needed in order to formulate a conceptual framework that has a validity in the musicological perspective (much like the generative theory of tonal music [4]). This framework would then be put to test with recent findings in cognition, notably on the topic of coupled oscillators, and finally a computer model of LRC will be developed.

Such a model would be of use both in musicological analytical settings and as a tool for compositional help and/or improvisational situations. LRC would therefore be not only an observational concept, but also a working generative environment for music productivity.

4. REFERENCES

[1] Dahan, K., Emergence, enaction et propagation des dimensions temporelles dans les processus compositionnels, *Filigrane n°10 : Musique et Rythme*, ed. Delatour, Paris, 2010.

[2] Forte, A., *The Structure of Atonal Music*, Yale University Press, 1977.

[3] Large, E., Neurodynamics of music, In Riess Jones, M., et al. (eds.), *Springer Handbook of Auditory Research, Vol. 36*, Springer, New-York, 2010.

[4] Lerdahl, F., Jackendoff, R., *A Generative Theory of Tonal Music*, MIT Press, Cambridge, 1996.

[5] Ruwet, N., *Langage, Musique, Poésie*, Seuil, Paris, 1972.

[6] Schenker, H, *Der freie Satz*, 1935.

[7] Xenakis, I., *Formalized Music*, Pendragon Press, New-York, 2nd ed. 2001; 1st ed. 1971.

Entrainment and pulse in groove-based music

Anne Danielsen and Alexander Refsum Jensenius

Department of Musicology, University of Oslo, Norway
anne.danielsen@imv.uio.no

Keywords: pulse, entrainment, dynamic attention, groove, motion capture

1. INTRODUCTION

For music in strict time, the basic pulse, also called the internal or subjective beat, is a fundamental reference for the production and perception of rhythm [1]. In the present paper, we investigate listeners' perceptual response to music where the shape of the internal beat changes during the musical course. We use as our example a tune by R&B artist D'Angelo entitled Left and Right. Previous analysis of the microrhythmic relationships of this tune shows that the first part of the tune implies an accurate and point-like location of the pulse, whereas in part two a cluster of pulse locations results in a much more vague internal beat [2]. First, we theorize how a confident listener may respond to the changes in the internal beat of the groove, using the theory of dynamic attending, and in particular its entrainment aspect, as the point of departure [4]. Then we report from a pilot study aimed at examining listeners' body movements in the different parts of the music by way of motion capture technology. Our hypothesis is that in part two of the tune listeners will be more loosely synchronized with the internal beat and thus respond with more 'rounded', inaccurate movements.

2. METHOD

Subjects: 16 music students were recruited to the study (age: mean=23 SD=3. Female=5, male=11).
Task: The subjects were seated during the experiment, with hands resting on their knees and a motion capture marker placed on top of their forehead. They were instructed to move their upper body to the pulse of the music.
Stimulus: Three different musical excerpts were chosen, one from part one of the tune (A), one covering the transition from part one to part two (B), and one from part two (C). Each excerpt lasted approximately 30 seconds, and they were played back with two seconds of silence between the tracks.
Motion capture: Recordings were done with an infrared optical marker-based motion capture system (Qualisys Oqus 300), running at 100 Hz. Data were recorded and pre-processed in Qualisys Track Manager and later analyzed in Matlab using the MoCap Toolbox [5].

3. RESULTS

The results for the quantity of motion show that the subjects on average moved more for sound C than for sound A by a factor of 1.3 (SD=0.4). Plots of the marker position indicate that the movements were more 'rounded' in response to sound C than to sound A. Plots of the response to sound B (the transition) show that for some subjects there is a clear change in movement pattern after the transition. Estimating movement periodicities using windowed autocorrelation [3], we find that most subjects entrained to the basic pulse of the music (~90bpm) in the vertical and sagittal planes, and to half-tempo in the transverse plane.

4. DISCUSSION

Given the tight coupling between action and perception in musical rhythm [6], the difference in movement response between part one and two of the tune may be interpreted as reflecting a difference in perceptual response (looser synchronization and a widening of the attentional focus in part two) caused by the microrhythmic alterations in the groove.

5. REFERENCES

[1] Chernoff, J.M. 1979. *African Rhythm and African Sensibilities.* (Oxford)

[2] Danielsen, A. 2010. Here, There and Everywhere. Three Accounts of Pulse in D'Angelo's 'Left and Right'. *Musical Rhythm in the Age of Digital Reproduction,* Ed. A. Danielsen. (Ashgate)

[3] Eerola, T., Luck, G., Toiviainen, P. 2006. An investigation of pre-schoolers' corporeal synchronization with music. *Proc. of the 9th Int. Conf. on Music Perception & Cognition*, Bologna, 472–476.

[4] Large, E., Riess-Jones, M. 1999. The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review* 106/1, 119-159.

[5] Toiviainen, P., Burger, B. 2010. *Mocap toolbox manual v 1.2.2.* Technical report, University of Jyväskylä.

[6] Styns, F. et al. 2007. Walking on music. *Human Movement Science* 26(5), 769–785.

Speaker identification based on speech rhythm: The case of bilinguals

Volker Dellwo, Stephan Schmid, Adrian Leemann, Marie-José Kolly and Matthias Müller

Phonetics Laboratory, University of Zurich, Switzerland
volker.dellwo@uzh.ch

Keywords: speech rhythm, speaker idiosyncratic features

1. INTRODUCTION

Voices are highly individual. The present study investigated how temporal characteristics of speech can contribute to speaker individuality for the same speaker speaking in different languages. By now there is a large body of evidence showing that measures based on temporal characteristics of consonantal and vocalic interval durations show drastic within language variability that is to a high degree a result of between speaker variability ([1], [2], [4]). Here we present results from an experiment on L2 and bilingual Italian/German speakers. Our assumption was that if speaker idiosyncratic rhythmic characteristics exist, then they should be present across utterances from different languages produced by the same speaker.

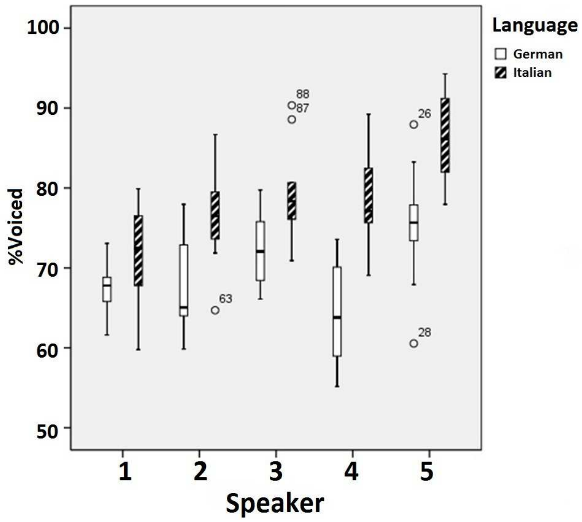
2. METHOD

5 Italian-German bilinguals, 5 German natives with L2 Italian and 5 Italian natives with L2 German were recorded reading 10 sentences (about 15 syllables on average) in each Italian and German. Speakers were selected to be similarly competent in their L2 across the two native language conditions (as judged by the second author). Durational characteristics of consonantal and vocalic intervals (e.g. syllable rate, %V, VarcoC and V, PVI, etc.) as well as voiced and unvoiced intervals (e.g. %Voiced, VarcoVoiced) were calculated for each sentence.

3. RESULTS & DISCUSSION

Results in figure 1 show the distributions (box-plots) of %Voiced (percentage over which speech is voiced) for each bilingual, speaking either Italian (striped plots) or German (white plots). Apart from the German condition of speaker 5 it can be observed that %Voiced has the tendency to increase from the leftmost to the rightmost speaker in both languages.

Figure 1: Box plot showing %Voiced for 5 bilinguals speaking German (white plots) and Italian (striped plots).



A factorial ANOVA (language * speaker) revealed that there is no interaction between the factors (p=0.065; the relatively low p-value is assumed to be caused by the German of speaker 4). However, highly significant main effects were found for both language (p<.001) and speaker (p<.001). The effects could be replicated for other metrics (e.g. VarcoV or syllable rate) and for the L2 speakers. Our results support the view that rhythmic characteristics of speech can be speaker-specific, independent of the language used by the speaker.

4. REFERENCES

[1] Dellwo, V. and Koreman, J. (2008) *How speaker idiosyncratic is measurable speech rhythm?* Abstract presented at IAFPA 2008, Lausanne/Switzerland (http://www.hf.ntnu.no/isk/koreman/Publications/2008/IAFPA2008abstract_DellwoKoreman.pdf)

[2] Loukina, A., Kochanski, G., Rosner, B. and Keane, E. (2011) *Rhythm measures and dimensions of durational variation in speech.* In: J. Acoust. Soc. Am. (129,5), 3258–3270.

[3] Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., and Mattys, S. (2010) *How stable are acoustic metrics of contrastive speech rhythm?* In: Journal of the Acoustical Society of America (127,3), 1559–1569.

[4] Yoon, T.J. (2010) *Capturing inter-speaker invariance using statistical measures of speech rhythm.* In: Electronic proceedings of Speech Prosody, Chicago/IL, USA.

Gender differences in a retrospective duration judgement

[1]David A Ellis, [1]Simon Kyle, [2]Richard Wiseman and [1]Rob Jenkins

[1]School of Psychology, University of Glasgow; [2]School of Psychology, University of Hertfordshire
d.ellis@psy.gla.ac.uk

Keywords: time perception, seconds, gender, retrospective

1. INTRODUCTION

Despite the implications for understanding sex differences in memory and cognitive processes, the issue of gender differences in psychological time is far from settled [1]. Psychological time involves processes by which a person adapts to and represents temporal properties in order to synchronise external events. For example, while crossing a busy street, speed and time estimates are continually required [2]. A good understanding of time is also vital when it comes to occupational and social organization. Because many everyday perceptual and cognitive situations lead a person to estimate short durations, it is important to understand the underlying processes and the existence of any individual differences. Researchers have previously explored this in small samples with results ranging from no gender effect to a significant female advantage in both retrospective and prospective duration judgments.

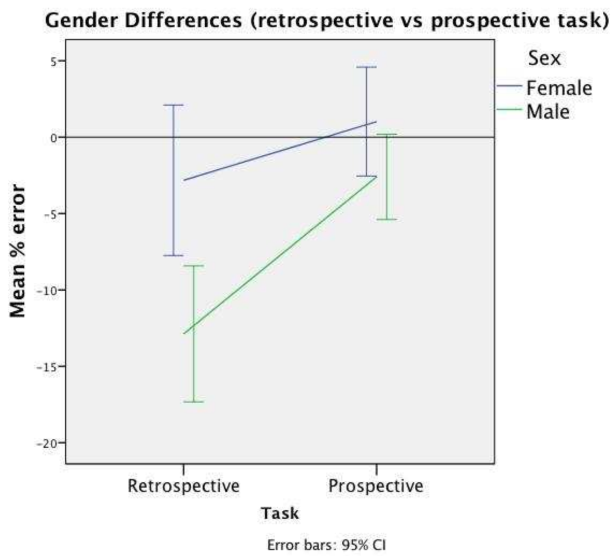
2. METHOD

In an online experiment, 526 (52.8% male) subjects completed two short duration estimation tasks. An audible 5-second tone was played while participants filled out a personality measure. Participants were unaware that this task was related to time perception and were prompted on the following page for a duration estimate in seconds (retrospective task). They were then instructed to estimate a second tone that lasted 7 seconds and provide another estimate (prospective task).

3. RESULTS

As expected, participants typically underestimate a retrospective duration and provide a more accurate estimate in the prospective task [$t(511) = -4.88$, $p<.0001$]. In addition, women demonstrated an improved accuracy in the retrospective task [$t(524) = -2.99$, $p<.0004$], but not the prospective task [$t(536) = -1.59$, $p = .112$].

Figure 1: Mean accuracy in the retrospective and prospective tasks between male and female subjects.



4. DISCUSSION

This study involved two very brief duration judgements, but the large sample supports previous findings [1]. To the best of our knowledge, this is the largest sample to test for these gender differences across a culturally diverse sample. A simple measure of personality did not reveal any significant correlations, and age was also not a factor. Evidence that females perform relatively better on episodic memory tasks than do males suggests that females are better at remembering events and therefore judge the retrospective duration as being longer [3]. It has also been suggested that, relative to males, females use different cognitive processes, especially as the delay between exposure and response increases [1]. Regardless of the possible other explanations, these findings are important for theories on gender differences in time perception.

5. REFERENCES

[1] Block, R. A., Hancock, P. A. and Zakay, D. (2000). Sex differences in duration judgments: A meta-analytic review. *Memory and Cognition*. 28 (8), 1333-1346

[2] Wittmann, M. and Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends in Cognitive Science*. 12 (1), 7-12.

[3] Herlitz A., Nilsson, L. G. and Backman, L. (1997). Gender differences in episodic memory. *Memory and Cognition*. 25, 801-811.

Regular timing patterns, sensorimotor synchronization and text-change detection in speech

Simone Falk[1]*, Simone Dalla Bella*[2,3,4]

[1]German Linguistics, Ludwig-Maximilians-Universität, Munich, Germany; [2]EuroMov, Movement to Health Laboratory, Université Montpellier I, Montpellier, France;[3]WSFiZ,Warsaw, Poland; [4]BRAMS, Montreal, Canada
simone.falk@germanistik.uni-muenchen.de

Keywords: text-change detection, motor synchronization

In the present study we examine the possibility that motor entrainment to speech affects memory processes in discourse comprehension. Predictable timing patterns (e.g., rhythmic music) are typically conducive to motor entrainment, as shown in sensorimotor synchronization experiments [e.g. 2]. The idea that entrainment also plays a role in speech has been raised [1] and recent findings show that motor synchronization to speech is possible and language-specific [3].

We argue that motor synchronization with predictable timing patterns in speech enhances the salience of relevant discourse entities (e.g. verbs). Heightened attention to these elements is expected to positively affect memory and encoding of sentence meaning during comprehension. This possibility is consistent with previous evidence of enhanced memory for discourse entities when speech is accompanied by iconic gestures [7]. Even simple beat gestures accompanying the speech signal seem to have a positive effect on memory processes [10]. Still, the link between motor synchronization and text comprehension has not been examined so far.

In this study, we tested 64 participants on a combined sensorimotor synchronization task (finger tapping; [4]) and memory task (i.e., the text-change paradigm; [6]). In the tapping task, participants were asked to tap their index finger in synchrony with a metronome (IOI = 600 ms) followed by a speech stimulus (in German). The stimulus consisted of two transitive sentences, each formed by eight regular occurring iambic feet (Interval between accented vowels = approx. 600 ms). When the metronome stopped, participants were asked to continue the tapping at the pace of the metronome while listening to the speech stimulus. The stimulus was presented in such a way that the tapping continuation either coincided with the accented syllables of the signal (congruent condition) or not (incongruent condition). The tapping stopped after the speech stimulus has been presented. It was presented again after a short pause, and in most of the cases, the second presentation contained a text change (i.e., a word has been changed). Participants had to indicate as quickly as possible if they perceived a change in the second presentation, as compared to the first, and indicate the element which was changed.

Previous studies [6] have shown that not all changes are perceived equally well; for example, a change involving semantically close words is typically more difficult to detect. We hypothesized that close semantic changes of verbs are easier to detect when participants are tapping congruently to the rhythm of the speech stimulus than when they are tapping incongruently. Our results are in keeping with this hypothesis, thus suggesting a positive role of motor entrainment in speech comprehension.

REFERENCES

[1] Cummins, F. 2008. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics* 37, 16-28.

[2] Large, E. W., Jones M.R. 1999. The dynamics of attending: how people track time-varying events. *Psychological Review* 106 (1), 119-159

[3] Lidji, P., Palmer, C., Peretz, I. & Morningstar, M. 2011. Listeners hear the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin and Review (published online)*.

[4] Repp, B.H. 2005. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, 12(6), 969-92.

[5] So, W.C, Chen-Hui, C.S., Wei-Shan, C. 2011. Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language & Cognitive Processes (published online).*

[6] Sturt, P., Sanford, A. J., Stewart, A., Dawydiak, E. 2004. Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review* 11, 882-888.

[7] Yap, D. F., So, W. C., Yap, M. J., Tan, Y. Q., Teoh R.L.S. 2011: Iconic gestures prime words. *Cognitive Science 35*, 171–183.

Convergence of speech rate: Interactive Alignment beyond representation

Ian R. Finlayson[1], Robin J. Lickley[1], Martin Corley[2]

[1]Clinical Audiology, Speech & Language Research Centre, Queen Margaret University, UK
[2]Department of Psychology, University of Edinburgh, UK
ifinlayson@qmu.ac.uk

Keywords: alignment, corpus study, dialogue, speech rate

1. BACKGROUND

It has long been known that conversational partners tend to align on common ways of talking about the world, not only in choice of syntactic structures or referring expressions [4], but also in manner [1]. Alignment in both of these areas has received considerable theoretical attention, however while accounts of the alignment of representations have considered the linguistic mechanisms responsible (for example the *Interactive Alignment account* [4]), theories of the alignment of performative aspects of conversation, such as speech rate, have largely tended to limit their scope to motivational explanations (most notably in *Accommodation Theory* [2]*)*.

One exception to this trend has been Wilson & Wilson's [6] oscillator model of turn-taking. In order to explain high coordination in turn-taking they propose that endogenous oscillators in the brains of conversational partners, representing their readiness to speak, have their frequencies determined by each other's speech rate. As these oscillators become entrained (as partners align on speech rate) this coordination should result in more seamless turn-taking. A crucial prediction of this model is therefore that as interlocutors' rates converge, the amount of variance in their turn-intervals (TI) should decrease. We investigate this by first establishing that speakers in dialogue converge on rate through local priming [3], and then testing if increasing convergence reduces variance in interlocutors' TI.

2. ANALYSES AND DISCUSSION

We established the articulation rate (AR) in 20,974 conversational turns from the Map Task Corpus [1], as well as the TI for each. As a measure of the amount of variance in TI we mean-centered, and then took the absolute value of, each turn-interval. With this manipulation, the mean should tend towards zero as variance decreases.

Using linear mixed-effects regression we investigated AR convergence and its relationship with turn-taking. Firstly, we modelled each speaker's AR as a function of their partner's AR in the previous turn. Secondly, we modelled the difference between interlocutors' rates in each turn across the length of the dialogue. Finally, we tested for a relationship between the differences in AR and our measure of TI variance.

Speakers' AR was found to be higher when their partner had spoken faster in the previous turn ($p < .001$), suggesting local priming of rate. This repeated, reciprocal, priming appeared to lead to convergence, with the absolute difference between interlocutors' rates decreasing throughout the duration of the dialogue ($p < .05$). No evidence was found of a relationship between the extent of convergence of rate and the amount of variance in TI ($p = .12$).

In absence of support for this important claim of the oscillator model we suggest that the interactive alignment account may extend beyond *what* is said, to *how* it is said. Specifically, we propose an account where the alignment of rate comes as a consequence of the use of production systems during comprehension [6]. Finally, our analyses provide a demonstration of the strength of sophisticated modelling techniques for investigating fine-grained aspects of timing within dialogue.

3. REFERENCES

[1] Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty, G. M., et al. (1991). The HCRC Map Task Corpus. *Lang. Speech*, 34:351–366.

[2] Giles, H., Coupland, J., and Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge Univ Pr, New York, NY.

[3] Jungers, M. and Hupp, J. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Lang. Cogn. Process.*, 24:611–624.

[4] Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.*, 27:169–190.

[5] Pickering, M. J. and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.*, 11:105–110.

[6] Wilson, M. and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.*, 12:957–968.

Temporal and melodic patterns of speech: Semantics and idiosyncrasy

Anatole Fiodorov

Department of English Phonetics, State Linguistics University, Minsk, Belarus
lotana@anitex.by

Keywords: rhythm, melody, syntagm, segmentation

1. INTRODUCTION

The interest to the temporal organization of speech is not fading. It plays a major role in the shaping of speech rhythm. Any distortion of temporal structure leads to disruption of speech and its comprehension. Temporal boundaries occurring at "wrong places" or odd durations of certain segments of speech chain may cause noticeable speech disfluencies.

Brigitte Keller et al [2] made an attempt to improve the synthetic speech quality. They identified a durational anticorrelation component that manifested itself within 500 ms, i.e. one or two syllables. They found that if a syllable tended to be short, the following syllable would be longer. The anticorrelation was .234, and they noticed positive perceptual effects. In the present study we attempted to show that the picture is a bit more sophisticated.

2. METHOD

In the present study we used speech material produced by two native speakers of English, professional linguists. They delivered three lecturettes on new words usage. The speakers were the authors of the texts which were read in an informal manner. Initially the recordings were transcribed and phonetically marked up. Then the recordings were segmented manually into words and measured using audio feedback and spectrograms with the help of PRAAT program. If a sound stretch didn't lend itself to accurate segmentation it was taken as one unit.

The duration of segments was calculated in ms without pauses and false starts. The average duration of a syllable was worked out as the mean proportional for the entire lecturette which approached the modal value. After that the average syllable duration was found for every word on the basis of its standard transcription. Syllable absolute durations were then expressed in relative values as to the total mean duration (mode) for a particular speaker.

3. RESULTS

The obtained results were transformed into graphs of the texts with a reference line at the origin at 1.0, axis Y. Similarly, the relative values of F0 were obtained with respect to the mode value of F0 for both speakers. The speech rate was measured as the average duration of a syllable in ms.

4. DISCUSSION

The analysis of graphs showed that the lengthening of syllables is associated primarily with the relative prominence of particular words in the speech flow. The data lead to believe that the speakers masterly control even negligible changes in relative syllable duration to lend a word an extra emphasis.

Stressed words are regularly signalled by above-the-average values of syllable relative durations. The values of relative durations make up consecutive series that are characterized by an increasing length of syllables. These series seem to be closely associated with syntagms, thus facilitating the identification of their boundaries. The findings strongly indicate that the syntagm is marked by a gradual slow-down of tempo, followed by other sequences.

In comparison with the melodic configuration of an utterance, temporal pattern proves to better reflect its semantic structure. If the melody curve is highly unpredictable and idiosyncratic, the temporal pattern looks more rigid and prescribed.

The suggested method makes it possible to pinpoint the boundaries of syntagms more accurately and allows for formalization of this procedure. There is no significant correlation between the distribution of syllable relative durations and F0 pattern.

5. REFERENCES

[1] Nolan, F. 2006. Intonation *The Handbook of English Linguistics,* (ed. Aarts and McMahon), Hardcastle, Blackwell Publishing Ltd, 433-457.

[2] Keller, B. 2002. Revisiting the Status of Speech Rhythm, *Speech Prosody 2002.* Aix-en-Provence, April 11-13, 2002, 727.

Mechanisms for the timely coordination of utterances

Chiara Gambi[1], *Joris Van de Cavey*[2], *and Martin J. Pickering*[3]

[1,3]Department of Psychology, University of Edinburgh, Scotland [2] Department of Experimental Psychology, University of Ghent, Belgium
[1]c.gambi@sms.ed.ac.uk

Keywords: dialogue, coordination, shared representations, prediction.

1. INTRODUCTION

Listeners are remarkably good at predicting when the current speaker's turn is going to end. De Ruiter et al. [1] suggested that this ability could explain why inter-turn intervals are mostly very short.

There is indeed evidence that predictive representations are implicated in language comprehension. Pickering and Garrod [4] proposed that predictions are computed *via* forward models in the listener's production system.

We extend their proposal and describe a set of mechanisms that could underlie the timely coordination of utterances in conversations [2].

2. THE MODEL

The core assumption of our account is that self-generated and other-generated utterances are represented in a similar format in the production system of a language user. In particular, forward models are computed for both types of utterances.

These are representations of aspects of the content, and timing, of the utterances, and are ready before the utterances are actually formulated. For this reason, they can be used to guide the process of utterance formulation. By taking such predictions into account, speakers can coordinate both subsequent fragments of their own utterance and their own and another's utterances [2].

Here, we focus on timing predictions. These are estimates of the time it takes to complete the various stages of formulation (syntactic encoding, lexical retrieval, phonological encoding) and, ultimately, to initiate articulation. Speakers are sensitive to timing estimates and use them to fine-tune the production of their own utterances (e.g., [3])

We propose that interlocutors make use of these estimates to coordinate with each other; in particular, to decide when to stop speaking (if they anticipate an interruption) and when to start speaking (if they anticipate their partner's turn is close to an end).

3. EXPERIMENTAL EVIDENCE

Twenty-four pairs of speakers described simple events (e.g., a nun following a doctor). They either produced a full sentence or stopped at the verb (e.g., *the nun follows*…), and let their partner continue.

The length of the completion varied (short vs. long; e.g., *the doctor* vs. *the doctor with the cane and the vase*). Participants who overlapped with their partner on more than 10% of the trials were discarded (N=11).

Speakers took longer from speech onset to verb offset when they were planning longer completions (long: 1006ms, short: 961ms; β=38ms, SE=17ms, t=2.28). Importantly, they also took longer when their partner was planning a longer completion (long: 1091ms, short: 1054ms), albeit only marginally so (β=29ms, SE=17ms, t=1.72). We argue they are sensitive to the fact that their partner needs more time to formulate a longer completion.

4. SUMMARY

We propose that timely coordination of the utterances of two speakers is based on the computation of timing predictions that guide the formulation and articulation of utterances.

5. REFERENCES

[1] De Ruiter, J., Mitterer, H., Enfield, N. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82, 515-535.

[2] Gambi, C., Pickering, M. 2011. A cognitive architecture for the coordination of utterances. *Front. Psychology* 2:275. doi: 10.3389/fpsyg.2011.00275

[3] Griffin, Z. 2003. A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review* 10, 603-609.

[4] Pickering, M., Garrod, S. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11**,** 105-110.

The influence of musical rhythm on the perception of word stress in Williams Syndrome

Reyna L. Gordon[1], *Miriam D. Lense*[1], *Cyrille L. Magne*[2], *Alexandra P.F. Key*[1], *Elisabeth M. Dykens*[1]

[1]Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, USA;
[2]Psychology Department and PhD in Literacy Studies, Middle Tennessee State University, USA
Reyna.Gordon@Vanderbilt.edu

Keywords: Williams Syndrome, stress patterns, EEG, domain-general, rhythm

1. INTRODUCTION

Williams Syndrome (WS) is a rare, neurodevelopmental genetic disorder associated with a unique behavioral phenotype including cognitive impairment, hypersociability, and musicality. Though verbal skills are typically superior to nonverbal skills, early language learning is delayed and there is evidence for atypical prosody. The aim of the present study was to determine if musical rhythmic context could influence processing of words in individuals with and without WS.

2. METHOD

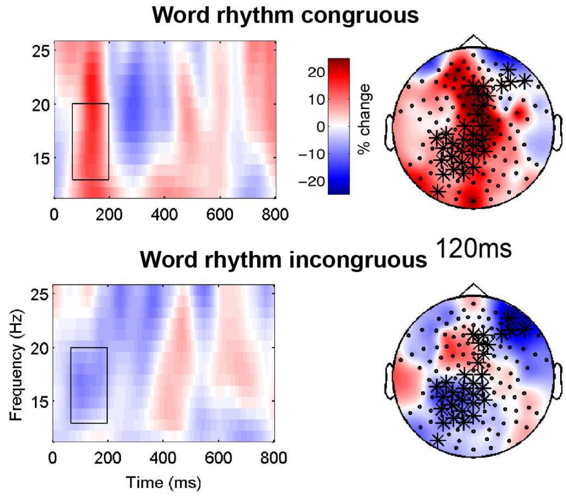
Sixteen young adults with WS (27.1±6.8 years, 6 females, IQ: 72.1±15.5) participated in this experiment. EEG was recorded while participants listened to sequences of instrumental tones interspersed with trochaic words (strong-weak stress pattern). The rhythm of the preceding tones was either congruous (long-short) or incongruous (short-long) with the trochaic words. Participants performed a timbre detection task on the instrumental tones and thus were not attending directly to the words.

Time-frequency analyses of EEG data were carried out on single trials time-locked to the onsets of the words, and averaged together, using the Fieldtrip toolbox. Cluster-based randomizations were carried out to identify significant differences in scalp topography of the resulting time-frequency representations.

3. RESULTS

Preliminary results reveal a cluster of electrodes (cluster p=0.062) showing a burst of increased induced beta power (13-20 Hz) between 64 and 196 ms post-word-onset for *word rhythm congruous* vs. *word rhythm incongruous*.

Figure 1: Time-frequency representations (left) and topographic plots (right, at 120ms) of grand average normalized induced beta band power in the two experimental conditions. Electrodes belonging to the cluster are marked with *, and the time and frequency of the cluster is delimited with the black box. The same scale applies to all of the plots.



4. DISCUSSION

These results show that musical rhythm expectations influence early implicit processing of spoken words in WS, and are consistent with domain-general rhythm processing mechanisms predicted by Dynamic attending theory [2].

These findings coincide with recent results showing beta band activity in mediating metrical expectations in instrumental and sung music [1], and that linguistic rhythmic context influences perception of stress patterns in spoken words [3]. Comparisons with typically developing participants are underway and will be presented along with behavioral assessments of rhythmic abilities in both groups.

5. REFERENCES

[1] Gordon, R.L., Magne, C.L., and Large, E.W. (2011). EEG Correlates of Song Prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology* 2.

[2] Large, E.W., and Jones, M.R. (1999). The dynamics of attending: How we track time varying events. *Psychol. Rev.* 106**,** 119-159.

[3] Magne C, Jordan D, Gordon RL (2011) Influence of Music Aptitude on Metrical Expectancy during Speech Perception. Third Annual Neurobiology of Language Conference. Annapolis, MD.

Chunking versus beat perception in auditory short-term memory

Jessica A. Grahn

Brain and Mind Institute & Department of Psychology, University of Western Ontario, Canada
jgrahn@uwo.ca

Keywords: rhythm, chunking, beat, fMRI, working memory

1. INTRODUCTION

Auditory working memory is often conceived of as a unitary capacity: working memory for different types of auditory materials (syllables, pitches, rhythms) is often thought to rely on similar neural mechanisms. The most influential model of auditory working memory is Baddeley's phonological loop model [1]. This model is largely developed on the basis of studies that use linguistic and verbal material, but perhaps could account for auditory working memory of other material, such as rhythm.

Verbal working memory studies indicate that 'chunking' is a spontaneous behaviour that benefits working memory performance [2]. For example, verbally recounted digit strings (such as those reproduced in digit span tasks) are often grouped into 3 or 4 numbers at a time, even when no grouping of the numbers is present during encoding. Perhaps analogously, temporal sequences with a regular beat structure (such as in music) are more accurately reproduced than irregular sequences. It is possible that beat-based structure enables encoding of rhythmic patterns in chunks, and that the same mechanisms underlying chunking in verbal working memory are at work in memory for beat-based rhythms. The current fMRI study examines this possibility, measuring brain responses to chunked and unchunked verbal sequences compare to beat-based and nonbeat-based rhythmic sequences.

2. METHOD

During fMRI scanning, participants performed rhythmic and verbal auditory discrimination tasks. Each trial consisted of stimulus presentation, a variable length delay, then a second stimulus presentation that participants judged to be same or different. In the rhythm condition, sequences were constructed from a single letter, repeated with rhythmic timing. Participants judged whether the rhythm changed. Verbal sequences were constructed from multiple letters, presented serially, and participants judged whether the order of letters changed. In each condition the stimuli were presented as chunks (beat-based rhythms of 2 or 3 musical bars for the rhythmic condition, grouping of letters into two 2- or 4-unit chunks for the verbal condition) or not chunked (nonbeat-based rhythms, or irregular temporal grouping of letters for the verbal condition). Comparisons between encoding, maintenance, and discrimination were made for both rhythmic and verbal sequences

3. RESULTS

Rhythm stimuli compared to verbal stimuli activated basal ganglia, supplementary motor area, and anterior insula during stimulus presentation (and a subset of these areas during discrimination). Verbal stimuli compared to rhythm stimuli activated bilateral auditory cortices (and these areas plus parietal cortex, precuneus, and middle frontal gyri during discrimination). In the basal ganglia, there was a significant interaction between the rhythmic/verbal factor and the chunked/not chunked factor: Beat-based rhythms activated the basal ganglia more than nonbeat-based rhythms, but verbal chunked and not chunked conditions did not show any difference in basal ganglia activity.

4. DISCUSSION

Overall, there was much overlap between rhythm and verbal working memory networks. However, the significant interaction in the basal ganglia suggests that beat perception is not a case of chunking. This suggests a dissociation between beat processing and grouping mechanisms that warrants further exploration.

5. REFERENCES

[1] Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp. 47–89). New York: Academic Press.

[2] Henson R.N., Burgess N., Frith C.D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia,* 38:426-40.

Going the distance: Effects of distal and global timing on word segmentation

Christopher C. Heffner[1,2]; Laura C. Dilley[3,1,2]; J. Devin McAuley[2]

[1]Department of Linguistics and Germanic, Slavic, Asian, and African Languages, Michigan State University, USA; [2]Department of Psychology, Michigan State University, USA; [3]Department of Communicative Sciences and Disorders, Michigan State University, USA
heffner3@msu.edu

Keywords: speech rate, word segmentation, prosody, context effects, phonetics

1. INTRODUCTION

It is well-recognized that there are few invariant acoustic cues that signal where word boundaries are located in casual speech. Listeners instead combine a variety of cues to segment speech [2]. It was recently found that acoustic cues referred to as distal prosodic cues—timing and pitch cues temporally removed from a possible word boundary within a sentence—also affect word segmentation [1].

The present work investigated a possible role for timing cues that were located further than a sentence away from a word boundary ambiguity (i.e., global prosodic cues). Evidence for a potential effect of global prosodic cues comes from multiple sources; for example, global timing context can affect perception of the timing of pure tones [3].

2. METHOD

We built on recent studies that have used sentence fragments containing acoustically ambiguous cues to the existence of a word boundary before a possible function word [1]. An example of such a fragment is *John said he would obey a rebel*, where an acoustically reduced realization of the critical word *a* is capable of blending phonetically with the previous content word *obey*.

In the present experiment, participants were assigned to one of four global timing conditions. A single invariant rate condition (the 1.4i condition) served as a baseline; in this condition, participants were exposed to sentence fragments with a "distal duration multiplier" of 1.4, with duration 1.4 times the originally recorded duration. In each of three variable rate conditions, participants heard sentence fragments played at five different distal speech rates. For these three variable rate conditions, the mean distal duration multiplier was 1.4, 1.2, or 1.6 times the duration of the original spoken fragment. That is, the sentence fragment duration was, on average, the same as (1.4v condition), faster than (1.2v condition), or slower than (1.6v condition) the speech rate of the baseline condition, respectively.

The task for participants was to transcribe sentence fragments. Their transcriptions of certain acoustically-ambiguous sentence fragments were used to determine a "critical word report rate", which represented how often participants reported hearing a function word (and, thus, perceived a word boundary) within an acoustically-ambiguous region of a sentence fragment.

3. RESULTS

In line with previous studies, distal speech rate was found to affect word boundary placement, with slower distal speech rates leading to a lower critical word report rate. In addition, global speech rate modulated word boundary placement, with relatively fast global contexts (e.g., 1.2v) leading to higher critical word report rates than relatively slow global contexts (e.g., 1.6v). This was true both for the 1.4 distal rate shared by all conditions ($F(2,52) = 3.38$, $p = 0.042$) and for the distal rates shared between the 1.2v and 1.4v global context conditions ($F(1,28) = 5.47$, $p = 0.027$).

4. DISCUSSION

The results here echo studies of non-linguistic acoustic processing [3] in finding effects of global timing on word boundary placement. These findings have consequences for many theories of word segmentation (e.g., [2]), which cannot presently accommodate the effects of temporally-removed prosodic information on word segmentation.

5. REFERENCES

[1] Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*, 1664-1670.

[2] Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477-500.

[3] McAuley, J. D., & Miller, N. S. (2007). Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences. *Perception and Psychophysics, 69*, 709-718.

Neural entrainment to frequency- and amplitude-modulated sounds

Molly J. Henry and Jonas Obleser

Max Planck Research Group "Auditory Cognition"
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
henry@cbs.mpg.de

Keywords: Entrainment; neural oscillations; amplitude modulation; frequency modulation; dynamic attending

1. INTRODUCTION

Speech and music are characterized by slow delta (1 – 3 Hz) and theta (3 – 8 Hz) fluctuations in the frequency and amplitude domains. It has been suggested by, e.g., dynamic attending theory [1] that listeners rely on frequency (spectral) and amplitude (temporal) modulations to facilitate perception. The current study examined oscillatory brain responses to frequency- and amplitude-modulated sounds that potentially underlie facilitated auditory perception.

2. METHOD

Four participants passively listened to complex tones that were frequency modulated (FM) or amplitude modulated (AM) at a rate of 3 Hz while the electroencephalogram (EEG) was recorded. Listeners completed one FM session and one AM session; the order of sessions was counterbalanced. Phase locking [2] and power were estimated from time-frequency representations of the single-trial EEG signal. Analysis focused on oscillatory brain activity at the stimulus frequency (3 Hz) and the first and second harmonics (6 Hz and 9 Hz).

3. RESULTS

For both FM and AM stimuli, phase locking was observed at the stimulus frequency (3 Hz), consistent with neural entrainment to stimuli characterized by periodic fluctuations in either frequency or amplitude (Figure 1). Phase locking was also observed at the first harmonic of the stimulus frequency (6 Hz), but specifically for FM stimuli. With respect to power (Figure 2), both FM and AM stimuli were associated with an increase in alpha power (9–10 Hz) relative to a silent pre-stimulus baseline. The alpha increase was larger for AM stimuli than for FM stimuli. Moreover, for AM stimuli, alpha power increased linearly over time during stimulation. Alpha power was induced, i.e., not phase-locked to the stimulus, as indicated by phase locking values near zero in the alpha range for both FM and AM stimuli.

Figure 1: Phase locking values for 3-Hz FM (left) and AM (right) stimuli (a.u.). The middle panel shows an FM-AM ratio; negative values indicate higher phase locking for FM and positive values indicate higher phase locking for AM.

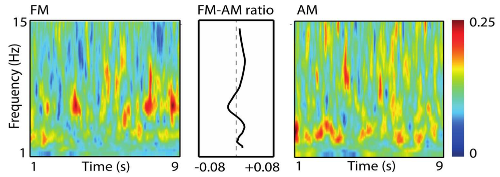
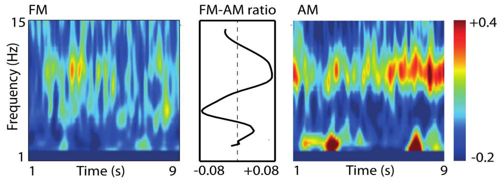


Figure 2: Power for 3-Hz FM (left) and AM (right) stimuli (proportional change from baseline). The middle panel shows an FM-AM ratio; negative values indicate higher power for FM and positive values indicate higher power for AM.



4. DISCUSSION

Slow ongoing brain oscillations entrain to both FM and AM stimuli at the stimulus frequency and to FM stimuli at the first harmonic frequency. This offers new avenues for studying the brain's excitation–inhibition cycle in simple auditory experiments. Although alpha enhancement is typically associated with inhibitory functions [3], alpha power increased during neural entrainment to both stimulus types relative to baseline, more so for AM stimuli. Further analyses will probe the relation between alpha power enhancement and entrainment.

5. REFERENCES

[1] Jones. (2003). Attention and Timing. In J. G. Neuhoff (Ed.), *Ecological Psychoacoustics*. New York. Academic Press.

[2] Lachaux, Rodriguez, Martinerie, and Varela. (1999) *Human Brain Mapping, 8*, 194–208.

[3] Weisz, Hartmann, Müller, Lorenz, & Obleser. (2011). *Frontiers in Psychology, 2,* 1-15.

Transmission of timing cues along a chain of individuals

J. J. Honisch, M. T. Elliott, A. E. Welchman, A. M. Wing

School of Psychology, University of Birmingham, England
m.t.elliott@bham.ac.uk

Keywords: sensorimotor synchronisation, motor timing, cue integration

1. INTRODUCTION

Often, we time and coordinate our movements as part of a group. This can take the form of an ensemble of musicians or group of dancers, for example. In these scenarios we may be presented with a number of timing cues simultaneously (e.g. in the case of a musical ensemble), possibly integrating the cues like those of a multisensory nature [2]. Alternatively, the timing cues may be passed along a group in the form of a chain (e.g. seen in some forms of dance, or a team of rowers [3]). In the latter situation, the passing of timing cues can be sequential, and hence is potentially subject to increasing variability, as each individual adds their own noise in terms of sensory processing and subsequent motor action.

Here, we investigate how the variability of movements is affected as timing cues are passed sequentially along a chain of people. In addition, we investigate how timing cues passed along two independent chains are integrated and synchronised with, when viewed by a single observer.

2. METHOD

Six participants (3 male, mean age: 37.2 years), sat on chairs in the formation shown in Figure 1. Participants produced oscillatory arm movements from the elbow. Movements were captured using motion capture cameras (Oqus, Qualisys Inc, Goteburg) via reflective markers attached to each participant's left and right index finger tip.

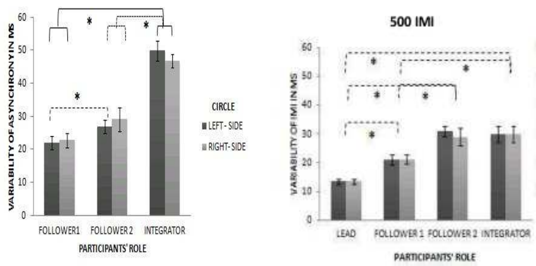
The lead (position 1), listened to an auditory metronome via headphones (using MatTAP [1]). A trial consisted of 30 slow (800 ms) and then 30 fast (500 ms) metronome intervals, or vice-versa. All other participants wore ear plugs to eliminate any external noise. Positions 2-5 were instructed to synchronise arm movements to that of the person to their right (2, 3) or left (4, 5). The integrator (6) was set slightly inward and synchronised their movements to those of both position 3 & 5. Participants rotated through each role during the experiment.

Figure 1: Group seating formation, creating two chains of timing cues: (2,3) and (4,5). Cues originate from (1) and are integrated by (6).

3. RESULTS

As predicted, both asynchrony and inter-response interval (IRI) variability increased in movements as cues progressed down the chain. However, while the integrator showed a substantial increase in asynchrony variability, IRI variability remained at the level of Follower 2.

Figure 2: Asynchrony variability (left) and IRI variability (right) of Lead (1), Follower 1 (2,4), Follower 2 (3,5) and Integrator (6).



4. DISCUSSION

Our results show that synchronised movements become increasingly variable as timing cues are passed along a chain of individuals. However, it appears the integrator does not attempt to synchronise to these resulting noisy cues, but instead extracts the underlying intervals. We will present a model of cue integration describing this behaviour.

5. REFERENCES

[1] Elliott MT, Welchman AE & Wing AM (2009). *J Neurosci Meth*. 117:250-257.

[2] Elliott MT, Wing AM & Welchman AE. (2010). *Europ J Neurosci*. 31:1828–1835.

[3] Wing AM & Woodburn C. (1995). *J Sports Sci*. 13:187–197.

Dynamic attending and the perceived duration of auditory oddballs

Elisa Kim and J.Devin McAuley

Department of Psychology, Michigan State University, U.S.A.
dmcauley@msu.edu

Keywords: attention, rhythm, timing, duration perception

1. INTRODUCTION

When an unexpected (oddball) stimulus is presented within a series of otherwise identical auditory or visual (standard) stimuli, the duration of the oddball stimulus tends to be overestimated. Explanations of the oddball effect have proposed that overestimation of the oddball duration is (a) due to an increase in attention to the unexpected oddball [1] or (b) conversely, an indirect consequence of habituation to the repeated standard [2]. The aims of the present study were to test both the enhanced attention and habituation hypotheses and also consider a dynamic attending account [3,4]. Specifically, we hypothesized that oddballs that occur earlier than expected would receive less attention and thereby be perceived as shorter than equivalent duration oddballs that occur at the expected time, while oddballs that occur later than expected would receive more attention and be perceived as longer than equivalent duration oddballs that occur at the expected time.

2. METHOD

Two experiments varied the timing of the onset of an embedded auditory oddball stimulus within an otherwise isochronous sequence of 8 standard stimuli. The serial position of the oddball (defined by a difference in pitch) varied randomly from trial to trial across positions 5, 6, 7, or 8. Listeners' task was to judge whether the duration of the oddball stimulus was 'shorter' or 'longer' than the duration of the standard. In Experiment 1, the sequence of standard stimuli established a regular (isochronous) rhythm at one of three rates (tempi) and the timing of the onset of the oddball was varied so that it was either 'early', 'on-time', or 'late' relative to an extrapolation of the isochronous rhythm established by the standard stimuli. In Experiment 2, the sequence of standard stimuli were given an irregular rhythm, but the same three fixed time intervals that preceded the onset of the oddball in Experiment 1 were examined.

3. RESULTS

Increasing the tempo of the standard sequence increased the perceived duration of the oddball. Consistent with a dynamic attending account, early oddballs tended to be underestimated, while late oddballs tended to be overestimated. On-time oddballs showed the smallest duration distortions. Irregular timed sequences weakened effects of oddball onset timing on perceived oddball duration.

4. DISCUSSION

Results showed that standard-sequence tempo and the relative timing of the onset of the oddball affect perceived oddball duration. Observed distortions in perceived oddball duration are not accounted for by either an enhanced attention hypothesis [1] or a habituation hypothesis [2]. Results are most consistent with a dynamic attending account [3,4] whereby early oddballs receive less attention and late oddballs receive more attention, relative to on-time oddballs.

5. REFERENCES

[1] Tse, P., Intriligator, J., Rivest, J., & Cavnaugh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics, 66*, 1171-1189.

[2] Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE 2(11): e1264*. doi:10.1371/journal.pone.0001264.

[3] Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review, 83(5), 323-355.

[4] Jones, M. R. & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459-491.

Rhythms of persuasion: The perception of periodicity in oratory

Sarah Knight and Ian Cross

Centre for Music and Science, University of Cambridge, UK
slk34@cam.ac.uk

Keywords: entrainment; speech; attention; social judgements

1. INTRODUCTION

Music and speech can be considered to be related as communicative modes through their shared capacity to establish joint attention and social bonds through real-time "relational" processes [1]. Inter-individual entrainment grounded in rhythmic regularity may be a key component of these processes: it powerfully channels attention [2] and promotes positive social attitudes [e.g. 3]. However, if rhythmic regularity is indeed important in bolstering communication's social efficacy, why is it not perceived in speech to the same degree as in music? We hypothesise that temporal regularity in speech is function-dependent; those speech registers which foreground relational processes – as opposed to information exchange – are perceived as more rhythmic and induce entrainment to a greater extent. This hypothesis was tested with a focus on persuasive oratory. Orators have a strong motivation to foreground relational processes, since to achieve their goal they must capture listeners' attention and foster positive attitudes to the same extent as, or to an even greater extent than, simply transmitting information.

2. STIMULI AND METHOD

16 speech excerpts of 19s were used (see Table 1).

Table 1: Stimulus information

Quantity and type	Comments
4 x conversational monologues	Generally agreed to lack periodicity.
4 x didactic speech (lectures)	Relies on capturing attention, but doesn't necessarily aim to induce favourable judgements of the speaker – hypothesised "halfway point" between conversation and oratory.
4 x persuasive oratory	Condition of interest.
4 x metrical poetry	Maximally rhythmic speech.

Each stimulus was played 3 times. While listening, the 22 participants were asked to tap along to "accents" or "stressed syllables" using a MIDI pad. Participants were also asked to rate each excerpt for "rhythmicity" after its final playing.

3. RESULTS

Participants rated the excerpts as follows, from lowest to highest perceived rhythmicity: conversation, didactic speech, oratory, poetry. This effect was highly significant overall ($F = 266.699$, $p < 0.0001$; Greenhouse-Geisser correction) and between all speech types ($p < 0.001$; post-hoc Bonferroni tests). Tapping variance also showed a highly significant effect of speech type, with variances ranging from highest to lowest as follows: conversation, didactic speech, oratory, poetry ($F = 24.001$, $p < 0.0001$; Greenhouse-Geisser correction). This effect was also highly significant between all speech types ($p < 0.001$; post-hoc Bonferroni tests).

4. DISCUSSION

The relatively low tapping variance and relatively high rhythmicity ratings for oratory imply that it is generally more rhythmic than both everyday speech and didactic speech. We suggest that this rhythmicity promotes relational processes, the outcomes of which further oratory's persuasive aim: that is, bouts of rhythmic regularity allow listeners to entrain to the speech, and this entrainment not only focuses listeners' attention but also promotes positive judgements of the speaker – both of which may serve to increase the speaker's persuasiveness. However, it is unclear whether rhythmic regularity actually serves these functions during speech perception. A study designed to investigate this question is currently being carried out, and preliminary results will be reported at the workshop.

5. REFERENCES

[1] Cross, I. 2011. Music and Biocultural Evolution. In M. Clayton, T. Herbert, R. Middleton (Eds.) *The Cultural Study of Music* (2nd edition) (London: Routledge)

[2] Barnes, R., Jones, M.R. 2000. Expectancy, Attention, and Time. *Cognitive Psychology* 41, 254-311.

[3] Hove, M.J., Risen, J.L. 2009. It's all in the timing: interpersonal synchrony increases affiliation. *Social Cognition* 27(6), 949-961.

Timing and prediction in audition: From sound to speech

Sonja A. Kotz

Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
kotz@cbs.mpg.de

Keywords: timing, rhythm, prediction, audition, dynamic attending

1. INTRODUCTION

Exposed to dynamic changes in the environment we need to trace the temporal structure of auditory (next to other) events to detect and adapt to these changes. While the duration of an event (i.e. sound) is considered as an inherent property, the regularity and order of perceived acoustic events define contextually extracted, statistically sampled temporal relations among events (Schwartze et al., 2011; 2012). These relations constitute the backbone of prediction in audition determining both "when" an event is likely occurring (regularity) and "what" type of event can be expected at a given point in time (order). In line with these assumptions, I present a novel cortico-subcortical neurofunctional model of temporal processing in audition that involves the division of labor between the cerebellum and the basal ganglia in the predictive tracing of acoustic events (Kotz et al., 2009; Kotz & Schwartze, 2010). Specifically, the cerebellum and its associated thalamo-cortical network appear to play a role in pre-attentive encoding of event-based temporal structure, while the attention-dependent basal ganglia-thalamo-frontal system is involved in the reanalysis and the re-sequencing of incongruent or unexpected temporal structure of a stimulus. I will discuss recent electrophysiological and fMRI data consistent with the proposed model and will bridge these data to speech perception and comprehension in healthy and patient populations (e.g. Kotz & Schwartze, 2011; Stahl et al., 2011).

2. REFERENCES

[1] Kotz, S.A., Schwartze, M. Schmidt-Kassow, M 2009. Non-motor basal ganglia functions: A review and proposal for a neurofunctional model of sensory predictability in auditory language perception. *Cortex,* 45(8), 982-990.

[2] Kotz, S.A., Schwartze, M. 2010. Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Science*, 14(9*)*, 392-399.

[3] Kotz, S.A., Schwartze, M. 2011. Differential input of the supplementary motor area to a dedicated temporal processing network: Functional and clinical implications. *Frontiers in Integrative Neuroscience,* 5, 86,doi:10.3389/fnint.2011.00086.

[4] Schwartze, M., Rothermich, K., Schmidt-Kassow, M., Kotz, S.A. 2011. Temporal regularity on pre-attentive and attentive processing of deviance. *Biological Psychology,* 87 (1), 146-151.

[5] Schwartze, M., Tavano, A., Schröger, E., Kotz, S.A. 2012. Temporal aspects of prediction in audition: Cortical and subcortial neural mechanisms. *International Journal of Psychophysiology*, 83(2), 200-207.

[6] Stahl, B., Kotz, S.A., Henseler, I., Turner, R., Geyer, S. 2011. Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*, 134, 3083-3093

Prosodic rhythms in reading and their contribution to text integration

Hamutal Kreiner

Linguistic Cognition Lab, Department of Behavioral Sciences, Ruppin Academic Center, Israel
Hamutalk@ruppin.ac.il

Keywords: prosody, rhythm, reading, text comprehension

1. INTRODUCTION

Prosody is manifested in the rhythm, intonation, and stress of speech and serves an important role in language comprehension. It has been shown to facilitate speech comprehension because it conveys pragmatic, syntactic and even lexical information. Furthermore, it has been argued that because prosodic structure corresponds to sentence structure it helps to organize and maintain information in working memory while the sentence is processed [1].

Like listeners, readers must be able to integrate information across several units to achieve sensible interpretation of the sentence as a whole, therefore they must maintain information in working memory. Here, we ask whether prosody facilitates integrative processing during reading as it does in spoken language? Unlike in speech comprehension, in reading, prosody is not part of the input. Yet, previous research shows that readers apply coherent prosodic rhythms that correspond to the structure of the sentence [4]. Moreover, it seems that implicit prosody applied in silent reading affects syntactic processing [3]. This study tested the hypothesis that disrupting the correspondence between prosodic rhythm and sentence structure disrupts integrative processing in reading. We test this hypothesis in two experiments: the first examines syntactic integration; the second examines semantic integration.

2. METHOD

In both Experiments participants read aloud sentences and their Eye-Movements (EM) were recorded. Half the sentences in Experiment 1 included violations of morpho-syntactic agreement (e.g., *the boys eats*) and half the sentences in Experiment 2 included semantically incongruent words (e.g., *the tables eat*). Previous EM studies show prolonged gaze-durations when readers encounter grammatical or semantic anomalies of these types [2]. This mismatching effect is assumed to be indicative of integrative processes. Finally, in both experiments participants were instructed to apply their natural reading rhythm to half of the sentences and monotonous rhythm based on grouping every 3 successive words in the other half.

3. RESULTS

Experiment 1 shows longer fixation durations on sentences that include violation of morpho-syntactic agreement compared to grammatical sentences. However, this mismatching effect occurs only when readers apply their natural reading rhythm; the effect is eliminated when monotonous grouped rhythm is applied. Experiment 2 shows late semantic mismatch effect when natural rhythm patterns are applied and this effect is further delayed when monotonous grouped rhythm is applied.

4. DISCUSSION

The results show that when monotonous grouped rhythm is applied by readers, syntactic integration is disrupted and semantic integration is delayed. These findings are consistent with our hypothesis that coherent reading prosody and in particular prosodic rhythm that corresponds to the structure of the sentence is involved integrative processing during reading. The proposal that natural prosodic rhythm facilitates incremental processing of the sentence because it conveys the structure of the sentence and helps maintaining it in the phonological loop in working memory is discussed in view of the results.

5. REFERENCES

[1] Cutler, A., Dahan, D., van Donselaar, W., 1997. Prosody in the comprehension of spoken language: A literature review. Language and speech, 40(2), 141-201.

[2] Deutsch, A., Bentin S., 2001. Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. Journal of memory and language, 45(2), 200-224.

[3] Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. Paper presented at the Speech Prosody, Aix-en-Provence, France.

[4] Koriat, A., Greenberg, S. N., Kreiner, H., 2002. The extraction of structure during reading: Evidence from reading prosody. Memory and cognition, 30(2), 270-280.

Resonating to rhythm

Edward Large

Centre for Complex Systems and Brain Sciences, Florida, USA
large@ccs.fau.edu

Keywords: rhythm, nonlinear resonance, cortical oscillation, dynamical systems

1. INTRODUCTION

I will explore the idea that certain classes of rhythms tap into brain dynamics at an appropriate time scale to cause the nervous system to resonate to rhythmic patterns. I will begin by reviewing behavioral studies that establish the role of rhythm in perceiving, attending to, and coordinating with rhythmic auditory stimuli [e.g., 1, 2]. I will then review neurophysiological and neuroimaging results from my lab and others, showing that cortical rhythms entrain to temporally structured acoustic signals [3], and entrainment of neuronal oscillations can function as a mechanism of attentional selection [4]. I will discuss the role of cortical oscillations in temporal expectancy [5], and the importance of subharmonic resonance [6]. I will cover amplitude and phase dynamics of cortical oscillations, and the role of delta, beta and gamma-band processes [7]. This discussion will include motor cortex involvement [8], cortico-striatal interactions [9], and possible mechanisms of auditory-motor interaction [10]. Finally, I will sketch a model intended to encompass rhythm perception, rhythmic attending, and rhythmic coordination. The approach is based on the idea that rhythmic expectancies emerge from neural resonance in sensory and motor cortices [11]. I will show that this approach may also have something to say about the communication of emotion in music [12].

2. REFERENCES

[1] Large, E. W., and Jones, M. R. 1999. The dynamics of attending: How we track time varying events. *Psychological Review*, *106* (*1*), 119-159.

[2] Loehr, J. D., Large, E. W., & Palmer, C. 2011. Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37 (4)*, 1292–1309.

[3] Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110-113.

[4] Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. 2010. Phase Entrainment of Human Delta Oscillations Can Mediate the Effects of Expectation on Reaction Speed. *Journal of Neuroscience, 30*, 13578-13585.

[5] Snyder, J. S., & Large, E. W. 2005. Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, *24*, 117-126.

[6] Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. 2011. Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*, 10234-10240.

[7] Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology, 94*, 1904-1911.

[8] Chen, J. L., Penhune, V. B., & Zatorre, R. J. 2008. Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex, 18*(12), 2844-2854.

[9] Grahn, J. A., & Brett, M. 2007. Rhythm perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893-906.

[10] Fujioka, T., Trainor, L. J., Large, E. W. & & Ross, B. 2012. Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *The Journal of Neuroscience*, *32*, 1791-1802.

[11] Large, E. W. 2008. Resonating to Musical Rhythm: Theory and Experiment. In S. Grondin, (Ed.) *The Psychology of Time*. West Yorkshire: Emerald.

[12] Chapin, H., Jantzen, K. J., Kelso, J. A. S., Steinberg, F. & Large, E. W. 2010. Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE 5(12),* e13812.

Rhythm processing in networks of neural oscillators

Edward Large

Centre for Complex Systems and Brain Sciences, Florida, USA
large@ccs.fau.edu

Keywords: neural networks, nonlinear oscillation, musical rhythm, syncopation

1. INTRODUCTION

The perception of musical rhythm is remarkable, in part because pulse and meter are perceived even in rhythmic stimuli that do not contain such periodicities. This phenomenon is most striking in syncopated rhythms, found in many genres of music, including music of non-Western cultures. In general, syncopated rhythms may have energy at frequencies that do not correspond to perceived pulse or meter, and perceived metrical frequencies that are weak or absent in the objective rhythmic stimulus. Here I propose a model based on the sketch developed in the first talk, 'Resonating to Rhythm.' The hypothesis is that the nonlinear processes that are observed in auditory physiology give rise to the remarkable features in human rhythm perception [1]. However, findings of nonlinear auditory processing challenge the assumptions of traditional perceptual models. I argue that the complex perceptual feats accomplished routinely by the human auditory system can be explained as the signal processing and pattern formation behaviors of canonical dynamical systems. More specifically, this approach models the auditory system using oscillatory neural networks that resonate to acoustic signals [2]. In this tutorial, I show how this approach can be cast into a mathematical theory using the tools of nonlinear dynamical systems. I will show how the resulting model predicts certain facts about behavior and neurophysiology [3].

2. REFERENCES

[1] Large, E. W. 2008. Resonating to Musical Rhythm: Theory and Experiment. In S. Grondin, (Ed.) *The Psychology of Time*. West Yorkshire: Emerald.

[2] Large, E. W., Almonte, F. & Velasco, M. 2010. A canonical model for gradient frequency neural networks. *Physica D, 239,* 905-911.

[3] Velasco, M. J. & Large, E. W. 2011. Pulse detection in syncopating rhythms using neural oscillators. *Proceedings of the 12th Annual Conference of the International Society for Music Information Retrieval*. (pp. 186-190).

Anti-phase synchronisation: Does 'error correction' really occur?

Jacques Launay, Roger T. Dean, Freya Bailes

MARCS Institute, University of Western Sydney, Australia
j.launay@uws.edu.au

Keywords: synchronisation; anti-phase; error correction

1. INTRODUCTION

Extensive psychophysical experimentation and modelling reveal that imperfections in our sensorimotor skills require complex monitoring and error correction processes in order to maintain accuracy when tapping to even an isochronous pulse [1]. While on-beat synchronisation with a stimulus has an obvious requirement for error correction, this is less clear with off-beat synchronisation, in which varying degrees of phase lock are possible in order to develop a steady relationship with isochronous stimuli. Yet error correction does occur in off-beat synchronisation, and has similar characteristics to those exhibited during on-beat tapping [2].

This study investigates whether error correction occurs in off-beat synchronisation when the instructions given oppose this behaviour, as has previously been shown with on-beat tapping [3]. Stimuli were anisochronous, with continuous deviations from an underlying pulse. The instruction to maintain a regular pulse required participants to ignore deviations, and not perform error correction.

2. METHOD

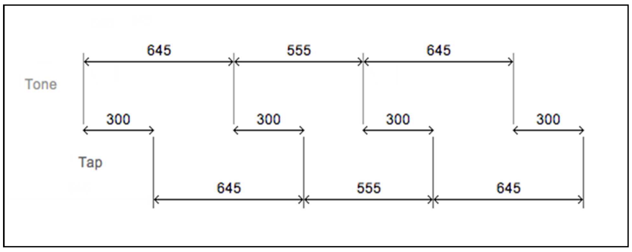
37 non-musicians were tested in a sensorimotor synchronisation paradigm, in which they listened to a sequence of slightly irregular sounds (devised similarly to [4]) played over headphones and were given the instruction not to synchronise with them. It was explained that they should "aim to tap at a different time from the tones", but try to maintain regularity and make one tap for each tone heard. Each trial lasted 96 taps, and the interonset interval of tones played was determined by one of two strategies employed by the computer system, either becoming increasingly isochronous or increasingly anisochronous.

3. RESULTS

Time series analysis, and modelling of individual data sets was performed in order to fully assess the data. Superficial analysis suggested that participants were adjusting their tapping in response to irregularities in the stimulus (with larger tap intervals following larger tone intervals). However, more complex modelling demonstrated that their behaviour is better explained by a variety of different strategies: maintaining a regular pulse, error correction, maintaining a regular asynchrony, or a form of negative error correction.

Figure 1 gives an example of how one of these strategies, maintaining a regular asynchrony, gives the appearance of responding to variability in an anisochronic sequence.

Figure 1: Maintaining a regular asynchrony with anisochronic tones gives the appearance of perfect error correction in the tap sequence. Vertical lines are events (taps or tones), and numbers are time between events (e.g. in ms).



4. DISCUSSION

The strategy employed by participants in this study cannot simply be described as error correction. However, the majority were not maintaining regularity of their own tapping (as instructed), but exhibiting changes in response based on the stimulus sequence. The concept of error correction during off-beat synchronisation needs some further exploration, as it is possible for people to perform in ways that are very different from on-beat synchronisation.

5. REFERENCES

[1] Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychon B Rev*, 12(6), 969-992.

[2] Chen, Y., Ding, M., & Kelso, S. J. (2001). Origins of timing errors in human sensorimotor coordination. *J Motor Behav*, 33(1), 3-8.

[3] Repp, B. H. (2002). Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *J Exp Psychol Human*, 28(2), 410-430.

[4] Madison, G., & Merker, B. (2004). Human sensorimotor tracking of continuous subliminal deviations from isochrony. *Neurosci Lett*, 370(1), 69-73.

Dialectal typology based on speech rhythm

Adrian Leemann, Volker Dellwo, Marie-José Kolly, Stephan Schmid

Phonetics Laboratory, University of Zurich, Switzerland
adrian.leemann@pholab.uzh.ch

Keywords: speech rhythm, Swiss German, rhythm metrics

1. INTRODUCTION

Studies on the segmental features of Swiss German (SG) dialects bear a long-standing tradition. In terms of speech rhythm as a parameter for prosodic typology of SG dialects, however, there is a clear lack of knowledge. In this contribution we examine whether or not SG dialects differ in terms of temporal characteristics and whether we can extract acoustic cues that account for these differences.

2. METHOD

To address this question we recorded 48 speakers of 8 different dialect regions (6 speakers per dialect): 2 Midland-Western Dialects Basel (BS) and Bern (BE), 2 Midland-Eastern dialects Thurgau (TG) and Zurich (ZH), 2 Alpine-Western dialects Sensebezirk (SB) and Valais (VS), and 2 Alpine-Eastern dialects Schwyz (SZ) and Grisons (GR). The following metrics were calculated: %V, ΔC, ΔV, varcoC, varcoV, rPVI-C, nPVI-C, and nPVI-V. For each rhythm metric we first ran a univariate ANOVA dialect by rhythm measure; second, we performed a t-test on Alpine/Midland dialects and thirdly, we conducted a t-test on Eastern/Western dialects.

3. RESULTS

The results of this study confirm distinct rhythmic differences between SG dialects for %V, vocalic, and consonantal variability. Vocalic variability seems to occupy a key role as shown particularly by the results of varcoV [1], see Figure 1.

Figure 1: Boxplots of the dialects' varcoV.

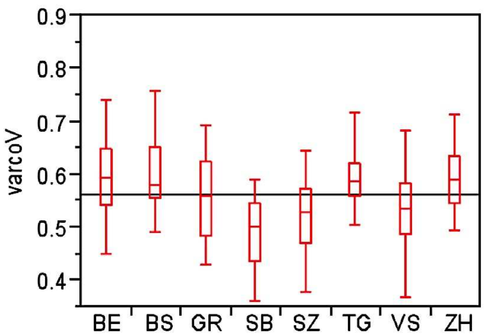


Figure 1 shows the boxplots of varcoV for each of the eight dialects. The horizontal line in the center of the figure illustrates the overall mean. The boxplots clearly suggest distinct differences between the dialects, which is confirmed by the ANOVA ($F(7, 184) = 8.2$, $p < .0001$); post-hoc tests corroborate significant variation. Taking into consideration also the results of %V and metrics of consonantal variability, we obtain a geolinguistic structuring of speech rhythm of the following order: Alpine-Western dialects demonstrate relatively little vocalic and consonantal variability; Midland-Eastern dialects show relatively high vocalic and consonantal variability; North-Western dialects as well as South-Eastern dialects to a greater or lesser extent fall in between these categories.

4. DISCUSSION

The current study showed that the variability of vocalic intervals occupies an important role for the discrimination of dialects – a result that has also been found by [2]. The findings further underline, however, that rhythmic variability is complex. Rhythm measures do not seem to vary uniformly across different dialects. With respect to the model of stress-timed and syllable-timed languages [3], we cannot detect such clear-cut categories in the current dialectal data. A significantly higher nPVI-C for Eastern as opposed to Western dialects, for instance, does not necessarily imply the same for varcoC. In other words, different rhythm metrics measure speech rhythm differently.

5. REFERENCES

[1] Dellwo, V., "Rhythm and Speech Rate: A Variation Coefficient for deltaC", in Karnowski, P., Szigeti, I. (eds.), *Language and language-processing*. Frankfurt am Main: Peter Lang, 231–241, 2006.

[2] Ferragne, E., and Pellegrino, F., "Rhythm in Read British English: Interdialect variability". *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju, Korea, 1573–1576, 2004.

[3] Pike, K. *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

L2 rhythm development by Mandarin Chinese learners of English

Aike Li and Brechtje Post

Department of Theoretical and Applied Linguistics, University of Cambridge, UK
al541@cam.ac.uk

Keywords: L2 rhythm, rhythm metrics, prosodic acquisition

1. INTRODUCTION

Traditional studies on linguistic rhythm attempted to classify languages as 'syllable-timed' vs. 'stress-timed' ([1], etc.). Although researchers have failed to find evidence for strict isochronicity in speech intervals, the perceived rhythmic differences that the classification tries to capture have been explained as emerging from distinct phonological and phonetic properties across languages, most notably syllable complexity and vowel reduction ([2], etc.).

Prieto *et al.* [3] examined the potential effects of a number of structural properties on rhythmic differences cross-linguistically, and found that although syllable structure does indeed have an effect, it cannot fully explain the data. In fact, an important source for the cross-linguistic differences in speech rhythm are the prosodic properties of accentual lengthening and phrase-final lengthening. Therefore, we can conclude that a range of systemic properties, including at least syllable complexity, vowel reduction, accentual lengthening and final lengthening, contribute to the characteristic rhythm of a language. The question arises how cross-linguistic differences in these properties may affect the acquisition of rhythm, and prosody in general, in a second language.

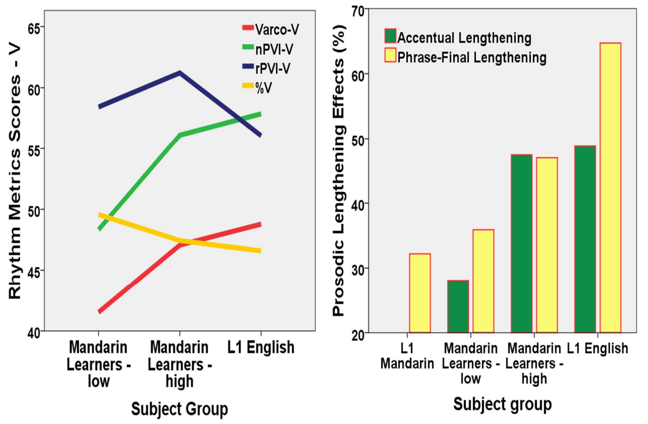
In this paper, we present a study of prosodic development in L2 English. Using a number of rhythm metrics (%V, Varco-V, Varco-C, Pairwise Variability Indices, cf. [4]) to quantify differences in rhythm, we investigated (1) to what extent rhythm metrics and the systemic properties which contribute to speech rhythm (accentual lengthening and phrase-final lengthening, here) reflect different levels of L2 proficiency, and (2) whether the development of these properties is reflected in L2 rhythm.

2. METHOD

The participants included two groups of L1 Beijing Mandarin learners of English with different English proficiency levels, and monolingual Beijing Mandarin speakers and native speakers of American English as control groups (5 per group). They read Mandarin Chinese or English sentences in which syllable structure was systematically varied. We analysed duration in syllables, vocalic/intervocalic intervals, and accented and phrase-final syllables.

3. RESULTS AND DISCUSSION

Our analyses show that the prosodic properties of accentual lengthening and phrase-final lengthening, as well as most rhythm metrics, discriminate well between proficiency levels; and they show an overall progression towards L1 English values. In addtion, the development of the prosodic properties is reflected in L2 rhythm.



These results may contribute to a multisystemic model of L2 rhythm development, in which the various systemic properties which contribute to speech rhythm interact with each other, and their development in the inter-language is reflected in the L2 rhythm acquisition process.

4. REFERENCES

[1] Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

[2] Dauer, R. M. (1987). Phonetic and Phonological Components of Language Rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences* , 447-450. Talinn.

[3] Prieto, P., Vanrell, M. d. M., Astruc, L., Payne, E. & Post, B. (2012). Phonotactic and Phrasal Properties of Speech Rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication 54*, 681-702.

[4] White, L. & Mattys, S. (2007). Calibrating Rhythm: First Language and Second Language Studies. *Journal of Phonetics*, *35*, 501-522.

Tapping rate affects tempo judgments for some listeners

Justin London, Emily Cogsdill

Carleton College, Northfield MN, USA
jlondon@carleton.edu

Keywords: tempo, tactus, tapping, kinematics

1. INTRODUCTION

Linkages between rhythm perception and motor behavior are well established in both behavioral [5] and neuroscientific [2][3][4] studies. [1] found that a number of musical parameters affected tempo perception and their effects can broadly be attributed to the amount of energy implicit in the acoustic signal. If higher levels of perceived energy lead to a faster tempo percept, then self-motion might also contribute to that perception. Or, more succinctly: the faster one taps, the faster the music will seem to move.

2. METHOD

A within-participants design was used; factors were tapping Mode (2) x Rhythmic Density (2) x Melodic Structure (2) x Tempo (6). Density (Dense vs. Sparse) was manipulated to contrast surface rate vs. beat rate, and melodic structure was manipulated to probe the effect of pitch motion. Two instances of each stimulus type were presented at six tempo levels: 72, 80, 89, 100, 120, or 133 bpm. Tapping mode (i.e., the rate of participant motion) had two levels: tapping every beat of the nominal tactus (E) or every other beat (EO). All melodies were in simple duple time.

Twenty-four participants heard all 48 stimuli twice and were organized into four blocks, with 24 stimuli in each block in a quasi-random order; presentation was counterbalanced amongst participants. After hearing each stimulus and tapping along with it participants marked their tempo rating on a tally sheet using a 9-point scale.

3. RESULTS

A repeated-measures ANOVA found no main effect for Tapping Mode ($F(1, 23) = 3.020$, $p = .096$), but did find significant main effects for Tempo ($F(2.016, 46.363) = 141.918$, $p < .001$), Rhythmic Density ($F(1, 23) = 37.429$, $p < .001$), and Melodic Structure ($F(1, 23) = 5.656$, $p = .026$; partial $\eta^2 = .197$; Greenhouse-Geisser correction applied). Two participants did not make consistent tempo ratings. For the remaining 22 participants, paired samples *t*-tests showed significant differences (two-tailed) in tapping mode (every beat vs. every other beat) for 8 participants, and non-significant results for 13 participants (4 of whom obtained *p* at or near one-tailed significance. Repeated measures ANOVAs for the "sensitive" subset of tappers now found (unsurprisingly) a main effect for Tapping Mode ($F(1, 7) = 49.657$, $p < .001$), as well as for Tempo ($F(1, 7) = 49.662$, $p < .001$), and Rhythmic Density ($F(1, 7) = 11.110$, $p = .013$), but no longer for Melodic Structure ($F(1, 7) = 4.109$, $p = .082$).

4. DISCUSSION

Overt sensorimotor engagement (i.e., tapping along) with a musical stimulus can affect its perceived tempo, at least for some listeners. We also found that percussive stimuli were perceived as moving faster than their melodic counterparts, as did [1]. The effects of tapping rate and melodic structure were only manifest at certain tempos, however, with sensitive tappers/listeners showing the effect of melody at a broader range of tempos than insensitive tappers.

5. REFERENCES

[1] Boltz, M.G. 2011. Illusory Tempo Changes Due to Musical Characteristics. *Music Perception* 28(4): 367-386.

[2] Chen, J. L., V. B. Penhune, and R. Zatorre. 2009. Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex* 18(12): 2844-2854.

[3] Grahn, J. A. and J. D. McAuley. 2009. Neural bases of individual differences in beat perception. *NeuroImage* 47(4): 1894-1903.

[4] Grondin, S. 2010. Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics* 72(3): 561-582.

[5] Repp, B. H. 2005. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review* 12(6): 969-992.

Application of rhythm metrics in disordered speech: What should we measure and what do they really tell us?

Anja Lowit

Speech and Language Therapy Unit, Strathclyde University
a.lowit@strath.ac.uk

Keywords: motor speech disorders, rhythm, variability, articulation

1. INTRODUCTION

It has long been established that speakers with motor speech disorders can present with rhythmic problems. However, quantifications of rhythmic disturbances have been relatively problematic, with early attempts not being generalisable to a wider range of tasks or reflecting the perceptual impressions of speech [1]. Investigations were taken further with the advent of acoustic rhythm metrics designed for cross-linguistic research. In particular, the PVI and VarcoV/C were highlighted as suitable to investigate the rhythmic properties of disordered speech [2].

Despite the success of the above rhythm metrics in differentiating disordered from unimpaired speech, or different disorders from each other, there have been few attempts to validate these measures against the perceptual evaluation of a speaker's rhythmic impairment. The only study to correlate perceptual ratings of speech rhythm with acoustic metrics has been by Henrich et al. [1]. However, this study only investigated the relationship between the measures in terms of severity and did not include qualitative assessment of the pattern of impairment. A further problem in relation to the clinical application of acoustic rhythm metrics is that there is insufficient normative data available, particularly in relation to performance variability across different speaking tasks.

The aim of this talk is to investigate how results of rhythmic metrics may vary across a variety of tasks in unimpaired speakers, and whether such performance variations are also observable in disordered speakers. In addition, the acoustic results will be compared to qualitative, perceptual analyses in order to establish the validity of the metrics in highlighting differences between patterns of impairment.

2. METHOD

The investigation draws on data from two separate studies. Study 1 investigated 10 speakers with ataxic dysarthria and 10 matched unimpaired speakers across a limerick, passage and sentence reading task and a monologue. Study 2 included three speakers with hypokinetic and three with ataxic dysarthria, and matched healthy controls performing a diadochokinetic and a sentence repetition task. A range of current acoustic rhythm metrics were applied to the data. In addition, speakers were rated perceptually for severity of rhythmic impairment and their speech was transcribed for qualitative analysis.

3. RESULTS & DISCUSSION

Results show that control speakers differ significantly across speech tasks, and that this performance is largely matched by the disordered speakers, with the exception of the monologue. Task choice is thus an important factor in any type of rhythm investigation. None of the metrics differentiated disordered from unimpaired speakers though, or disordered speakers from each other, despite clear perceptual differences. The phonetic analysis showed that different underlying deficits can result in similar rhythm scores, and that other factors beyond segment duration might impact on rhythm perception. Clinical evaluation of speakers thus needs to go beyond simple durational measures in order to capture the rhythmic disturbance experience by the speaker.

4. REFERENCES

[1] Henrich, J., Lowit, A., Schalling, E., & Mennen, I. (2006). Rhythmic Disturbance in Ataxic Dysarthria: A Comparison of Different Measures and Speech Tasks. *Journal of Medical Speech Language Pathology, 14*(4), 291-296.

[2] Liss, J. M., et al. (2009). Quantifying Speech Rhythm Abnormalities in the Dysarthrias. *Journal of Speech Language and Hearing Research, 52*(5), 1334-1352.

Music aptitude influences speech meter sensitivity

Cyrille Magne[1], Deanna Jordan[1], Reyna L. Gordon[2]

[1]Psychology Department & Program in Literacy Studies, Middle Tennessee State University, USA; [2]Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, USA
Cyrille.Magne@mtsu.edu

Keywords: music aptitude, speech, meter, ERP

1. INTRODUCTION

Speech meter, the pattern of stressed and unstressed syllables, is generally acknowledged to play a role in language in both infants and adults [3]. While meter is known to be an important aspect of speech perception, its neural basis has been largely understudied. The present study investigates the electrophysiological correlates of metrical expectancy in American English and the influence of music aptitude on speech meter sensitivity.

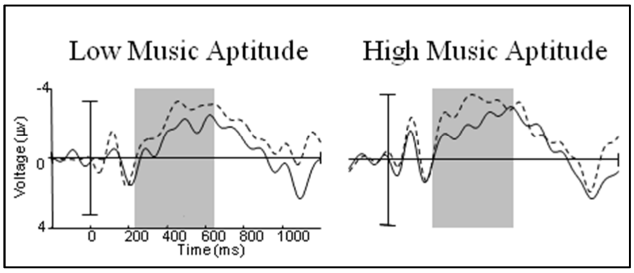
2. METHOD

ERPs were recorded in 18 participants while they were listening to 120 sequences of four bisyllabic words. The first three words were either all trochaic or all iambic while Metrical expectancy was manipulated by varying the stress pattern of the last word of each list, so that it had either the same or different stress pattern as the previous four words, while a pseudo-random inter-word interval was used to prevent temporal expectancy of successive stressed syllables. Participants were required to perform a memory task on each word sequence. Prior to the ERP session, the participants' music aptitude was assessed using the Advanced Measures of Music Audiation [2].

3. RESULTS

Fourth words that did not match the stress pattern of the previous words elicited significantly larger negative components between 250 and 650 ms following the word onset, over the fronto-central regions of the scalp. Individuals with high music aptitude displayed a larger negativity than individuals with low music aptitude over the left-hemisphere. In addition, for unexpected iambic words, significant positive correlations were also found between the amplitude of the negativity and the participants' score on the AMMA.

Figure 1: ERPs time-locked to the onset of metrically expected words (solid line) and metrically unexpected words (dashed line) are presented for the left centro-temporal region in participants with low music aptitude (left panel) and high music aptitude (right panel). The gray shaded area represents the 250-650ms latency range.



4. DISCUSSION

The results are in line with previous studies that similarly reported increased negativities in response to metrically/rhythmically incongruous or unexpected words in Dutch [1], French [4] and German [5]. Moreover, this negative ERP effect for metrically unexpected fourth words suggests that metrically regular word sequences induce expectancies about the stress pattern of upcoming words, even when the time-interval of the successive stressed syllables is irregular. Finally, our results suggest that music aptitude enhances speech meter sensitivity, particularly for the less common iambic stress pattern.

5. REFERENCES

[1] Böcker, K. B. et al. (1999). An ERP correlate of metrical stress in spoken word recognition. Psychophysiology, 36, 706-720.

[2] Gordon, E. E. (1989). Manual for the advanced measures of music audiation. Chicago, IL: G.I.A. Publications.

[3] Jusczyk, P.W. (1999). How infants begin to extract words from speech. Trends in Cognitive Sciences, 3, 323-328.

[4] Magne, C. et al. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. Cerebral Cortex, 17, 2659-2668.

[5] Schmidt-Kassow, M. and Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. Journal of Cognitive Neuroscience, 21, 1693-1708.

An oscillator based modeling of German spontaneous speech rhythm

Zofia Malisz[1], Benjamin Inden[2], Ipke Wachsmuth[2], Petra Wagner[1]

[1]Faculty of Linguistics and Literary Studies, [2] Faculty of Technology, Bielefeld University, Germany
zofia.malisz@uni-bielefeld.de

Keywords: rhythm, entrainment, oscillator

1. INTRODUCTION

Our approach assumes that speech rhythm is "manifested as the temporal binding of events to specific and predictable phases of a subordinate cycle" [1]. With this in mind, focus is laid on:

1) properties of the signal that represent basic periodic events. The vocalic pulse represents the basic timing coordination of the articulatory system on the "syllable-sized" level [2].

2) properties of the perceptual system entraining to abstract meter. The prominent vowel onset expresses the binding to the abstract subordinate foot cycle in German.

We use dynamically adaptive oscillators to entrain to the rhythmic properties in 1) and 2).

The approach accommodates concerns that interval duration analyses are not sufficient to provide evidence for perceived periodicities ("syllable-"or"stress-timing"). It also acknowledges evidence for hierarchical metric structures (e.g. metrical trees) to which the perceptual system resonates at multiple timescales. The above are mostly ignored by the "rhythm metrics" paradigm that focuses on static, interval timing. We propose to explore the dynamics of pulse timing at multiple timescales in speech rhythm analysis (also: [3],[4]).

2. METHOD

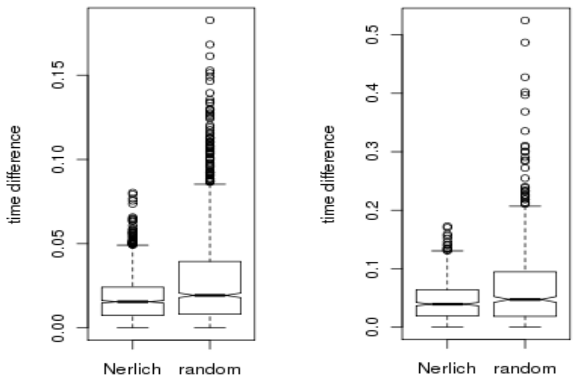
We first semi-automatically extract p-centres in the data [1],[2], i.e. vowel onsets; the basic pulse of the syllabic oscillator. Secondly, experts annotate rhythmic feet, representing the slower stress oscillator, where each prominent vowel is a pulse on that level.

The two levels are provided as input to several oscillator models simulating the spoken oscillations. The oscillatory models used are: adaptive oscillators developed for speech perception [5],[6] and music perception [3], general models and a neuronal model. We compare their performance to three control conditions: a random condition (corresponding to "blind guessing"), a "phase rule" condition with a phase reset and a "phase and period rule" condition where both phase and period are reset when input arrives. The periods of the oscillators are adjusted to the individual speaker mean within one recording session.

3. RESULTS

As shown in Fig. 1, the McAuley adaptive oscillator [5], modified by Nerlich [6] by an oscillation decay when input pulses are missing, performs best when entraining to fully spontaneous German speech data. The rule based conditions also outperform the random control (not shown).

Figure 1: The accuracy of oscillator entrainment to vowel onset based input (on the left) and foot based input (on the right). Only the oscillator with the best entrainment performance is shown.



4. DISCUSSION

An adaptive oscillator developed to model speech perception is able to entrain to rhythmically relevant speech input. Parameter optimisation by evolutionary algorithms and pulse frequency information as well as work on extracting periodic structures from the signal using acoustic methods [4],[7] is in progress.

5. REFERENCES

[1] Cummins, F., Port, R. 1998. Rhythmic constraints on stress timing in English. Journal of Phonetics 26, 145-171.

[2] Barbosa, P.A. 2006. Incursões em torno do ritmo da fala, Campinas: Pontes.

[3] Large, E. W. 2008. Resonating to musical rhythm: Theory and experiment. In: S. Grondin, (Ed.) The Psychology of Time. West Yorkshire: Emerald.

[4] Tilsen, S., Johnson K. 2008. Low-frequency Fourier analysis of speech rhythm. J. Acoust. Soc. Am. 124, 2.

[5] McAuley, J. D. 1995. Perception of time as phase. Phd dissertation, Indiana University, Bloomington.

[6] Nerlich U. 1998. Rhythmische Segmentierung sprachlicher Instruktionen in einem Mensch-Maschine-Kommunikations-Szenario. Unpublished M.Sc. thesis, Bielefeld University.

[7] Tamburini, F., Wagner, P. 2007. On automatic prominence detection for German. Proceedings of INTERSPEECH, Antwerp, 2007.

Tapping variability predicts rhythmic acuity

Fiona Manning, Michael Schutz

MAPLE lab, Psychology, Neuroscience and Behaviour, McMaster University, Canada
manninfc@mcmaster.ca

Keywords: tapping, timekeeping, perception and action, rhythm, sensorimotor integration

1. INTRODUCTION

Tapping has been used extensively in research to investigate synchronizing simple movements with auditory information [2]. Recent studies have demonstrated that movement can contribute to different aspects of auditory perception, such as pulse extraction [3]. We have previously demonstrated a relationship between movement and timing perception, where tapping along to an auditory beat can objectively improve the ability to detect timing deviations [1]. Here we further investigate this relationship by examining the relationship between tapping variability (consistency) and the benefits of tapping with respect to understanding a tone sequence's rhythmic structure.

2. METHOD

Each trial consisted of fourteen tones as shown in Figure 1. In the last group, the 2nd, 3rd and 4th tones were silent. The four groups were followed by a final probe tone. In half of the trials the probe tone was consistent with the pattern (offset = 0ms), and in the other half of the trials the probe tone occurred earlier or later than expected. In all trials, participants were instructed to identify whether the probe tone was consistent with the previous pattern.

In the experiment, participants were presented with a total of 64 trials; eight blocks each containing eight trials. Participants were asked to tap along to half of the blocks (movement condition) and asked to not tap during the other half of the blocks (no-movement condition).

During movmement blocks, participants tapped on a drum pad using a drumstick for all beats including the probe tone. Tapping information (including the timing and velocity of each tap) was recorded by the drumpad for analysis. During the no-movement blocks, participants were instructed to remain as still as possible.

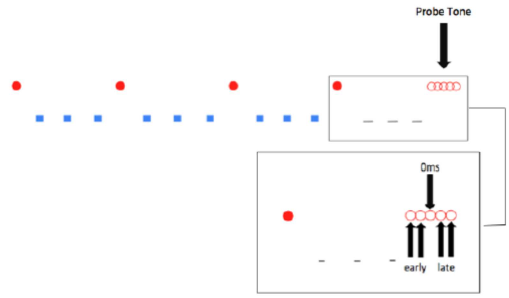


Figure 1: Example of a trial. The shapes indicate a sounded beat. The possible probe tone locations are labeled.

3. RESULTS

We calculated the effect of tapping on task performance by subtracting the proportion of correct responses in the no-movement trials from the proportion of correct responses in the movement trials. We analyzed tapping information by calculating the variance of beat size. In general, the data showed that tapping variability was significantly correlated with the effect of tapping, where the lower the tapping variability, the higher the effect of tapping.

4. DISCUSSION

This study demonstrates a relationship between tapping variability and the effect of tapping on performance (i.e., a relationship between tapping consistency and the ability to detect perturbations in the timing of the final probe tone). These data suggest that "better" tappers are more able to use movement as a means for keeping time. Overall these findings contribute to research on sensorimotor integration and literature on timing perception and action.

5. REFERENCES

[1] Manning, F. & Schutz, M. (2011, October). Moving to the beat improves timekeeping in a rhythm perception task. Paper presented at Acoustics Week in Canada. Québec City, Québec.

[2] Repp, B. H. 2005. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review* 12, 969-992.

[3] Su, Y.-H. & Pöppel, E. 2011. Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research*.

Modelling groove microtiming with pulse-coupled oscillators

Andy McGuiness

andymcguinesss@yahoo.com.au

Keywords: entrainment, microtiming deviations, groove, Funky Drummer

1. INTRODUCTION

The existence of systematic microtiming in groove music is a well-known phenomenon. The eight bar drum solo by Clyde Stubblefield on James Brown's *The Funky Drummer* [1] is typical of groove microtiming, with the first beat of each bar being longer than the other beats by a duration of tens of milliseconds [2]. This systematic variation in beat length within the bar is greater than the variation in bar duration across the eight bars of the drum solo.

2. METHOD

I present a computational model (implemented in Max/MSP) of two pulse-coupled oscillators. The oscillators are modelled on 'leaky integrate-and-fire' neurons, which fire at a given threshold: input resulting from the firing of one oscillator raises the current potential of the other oscillator, bringing it closer to its threshold in a shorter time. Onsets (16^{th} notes) between beats are considered to be under the control of a different timing mechanism.

The two oscillators have natural periods in the ratio 3:4. One oscillator (the 'overt' oscillator) reproduces the four beats per bar which are heard in the music. The other oscillator (the 'covert' oscillator) has three beats to the overt oscillator's four. The covert oscillator is considered to represent an activity of the drummer's brain which affects the timing of the the musical peformance but does not find overt expression in the performance.

3. RESULTS

The system of two pulse-coupled oscillators reproduces the characteristic pattern of long first beat, short second and third beat, and slightly-shortened fourth beat found in *The Funky Drummer*. The input from each oscillator to the other distorts the other's timing. The covert oscillator (with natural period 4/3 of the overt oscillator) distorts the timing of the overt oscillator so that it reproduces the characteristic /long-short-short-slightly shortened/ timing pattern found in *The Funky Drummer* break. By experiment, it was found that coupling strength from overt to covert oscillator needs to be approximately double coupling strength from covert to overt oscillator, to achieve the desired microtiming variation pattern. With natural frequencies set at a 4/3 ratio, the two oscillators entrain within several iterations (the exact number depends on their relative phase at start). Significantly, although the oscillator model is slightly noisy with regard to timing, the interaction of the two oscillators stabilises the bar duration to a variation of less than 15 milliseconds, similar to timing in *The Funky Drummer* break.

4. DISCUSSION

The notion of pulse-coupled oscillators as controlling musical beat timing generalises to a conception of human musical entrainment as involving populations of mutually pulse-coupled oscillators at graded frequencies. Periodic stimuli (such as are found in music) produce immediate and temporary phase alignment of many oscillators, but only those close to the period of the stimulus remain in phase [3][4]. A population of oscillators of similar natural frequencies recruited by a regular stimulus will remain in phase after the stimulus ceases, so underpinning the phenomenon of retention of beat tempo. Since pulse-coupling can only shorten the natural period of an oscillator, *accelerando* is possible utilising at least some of the same population of oscillators; while *rallentando* would correspond to the dropping out of some (faster) oscillators and the relaxing of others towards their natural frequency. *Ritardando* would involve a definite shift to a (mostly) different population of oscillators with longer natural periods.

5. REFERENCES

[1] Brown J., 1986, In The Jungle Groove, Polydor.

[2] Freeman P., and Lacey L., 2002, "Swing and groove: Contextual rhythmic nuance in live performance," 7th International Conference on Music Perception and Cognition, C. Stevens, D. Burnham, M. G, E.Schubert, and J. Renwick, eds., Sydney.

[3] Will U., 2005, "EEG Responses to Period Auditory Sitmuli and Brainwave Entrainment," Time, G. Turow, and J. Berger, eds., pp. 149-168.

[4] Nozaradan S., Peretz I., Missal M., and Mouraux A., 2011, "Tagging the neuronal entrainment to beat and meter.," The Journal of Neuroscience : the official journal of the Society for Neuroscience, **31**(28), pp. 10234-40.

Are we losing figural rhythm conceptions as future music teachers? A study based on J. Bamberger's graphic meaningful representations of rhythm produced by children and adults

Gabriele Menz

University of Education, Weingarten, Germany
menz@ph-weingarten.de

Keywords: rhythm, figural, Bamberger, notation, perception

1. INTRODUCTION

In the 1970s, J. Bamberger found a parallel development of figural and metric rhythm understanding in kindergarten and elementary-school students. Metric rhythm understanding is linked to durations of events mapped onto an underlying steady beat as in standard music notation. People with figural conceptions additionally account for context by emphasizing small-bound figural groups when listening to simple rhythms. Musical training supposedly produces some "wipe out" effect on figural understanding [1].

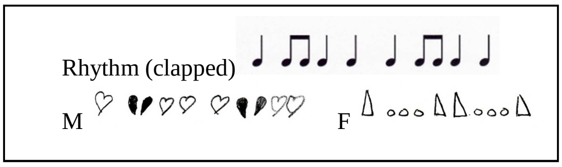
2. METHOD

Sixty-six undergraduate music education majors attending a US university took part in the study. The present part of the study tested their evaluation of children's meaningful graphic representations of simple rhythms and their own graphic representations of acoustic rhythm examples. Their drawings were classified by two independent raters according to Bamberger's categories. The study examined metric and figural classifications of participants' drawings and potential relationships between future music teachers' own graphic rhythm conceptions and their rating of kindergarten- and elementary-aged students' meaningful representations of rhythm.

3. RESULTS

Participants showed significantly more metric than figural inclinations in their own graphic conceptions. Yet, the number of participants evaluating in a predominantly metric way was found not to be significantly different from the number of evaluators who also evaluated children's figural drawings as meaningful representations of rhythm. Participants' drawing inclinations were found not to be independent from their evaluation proclivities. No associations were found between participants' categorizations as metric or figural drawers and the demographic variables gender, music education emphasis, or college level. Examples of metric (M) and grouping / figural (F) rhythm conceptions as drawn by participants are shown in Fig. 1.

Figure 1: Examples of a metric and a figural (grouping) graphic representation of rhythm 1 (only clapped).



4. DISCUSSION

Such an association between one's personal internalized conception when graphically representing simple rhythms and one's evaluation proclivity should be made known to future music teachers. In the present study, only half of the pre-service music teachers showed tolerance also toward meaningful figural drawings of children although children demonstrate perfect metrical competency when performing from their figural drawings [2]. Multiple hearings of rhythm should be fostered in the school system, and predominantly metric teachers should be aware of the potentially harmful effect of enforcing standard music notation on figural rhythm perceivers without mutual understanding. New musical rhythm standard tests should be designed for interdisciplinary research.

5. REFERENCES

[1] Bamberger, J. 1982. Revisiting children's drawings of simple rhythms: A function for reflection-in-action. In S. Strauss (Ed.), *U-shaped behavioral growth*, 191-226. New York: Academic Press.

[2] Smith, K. C., Cuddy, L. L., Upitis, R. 1994. Figural and metric understanding of rhythm. *Society for Research in Psychology of Music and Music Education 22*, 117–135.

The rhythm window

Oliver Niebuhr

Analysis of Spoken Language, Department of General Linguistics, ISFAS, University of Kiel, Germany
niebuhr@isfas.uni-kiel.de

Keywords: speech rate, prominence, emphasis, rhythm, German

1. INTRODUCTION

The presented study pursues the question whether the speaking-rate continuum that speakers can produce contains a rhythm window, i.e. a range of values that are particularly suitable for creating and identifying rhythmic structures. This question arose from the following observations. In investigating emphatic expressions in German, Niebuhr and colleagues found a type of 'emphasis for attention' whose essential characteristic is to split up utterances into their individual syllables and to create equal perceptual prominences on each syllable. Measuring the speaking rates of these utterances revealed a positive correlation between the number of syllables and the speaking rate with which they are produced. Given this finding, we analyzed the Kiel Corpus of Spontaneous Speech and found a similar correlation. However, this correlation between the number of syllables in an intonation phrase and the speaking rate of this phrase showed upper and lower thresholds of 7.9 and 3.1 syll/sec. The upper threshold is not due to biomechanical limitations, as speech production can be much faster on request. The lower threshold is well above 2.3 syll/sec, which was the upper threshold in the 'emphasis for attention' productions. So, speakers can be slower than 3.1 syll/sec. Against this backdrop, we assumed that the speaking rates used in spontaneous (fluent, non-emphatic) conversation are limited by the fact that the produced utterances comply with a rhythm window. Outside this window, prominence differences and hence rhythmic structures become flattened and/or less robust. Thus, e.g., the speaking rates of 'emphasis for attention' do not exceed 3 syll/sec in order to avoid creating a differentiated rhythm.

2. METHOD

We tested our assumption in a perception experiment based on reiterant [b̥a] syllables that we got from experiments of Kohler. Three different looped stimuli were used: (1) a sequence of alternating strong and weak [b̥a] syllables, (2) a sequence of alternating weak and strong [b̥a] syllables, and (3) a sequence of physically identical [b̥a] syllables. Weak and strong refer to prominence relations, with the strong syllables having longer durations, higher intensity levels, and higher F0 peaks. For each sequence, a speaking-rate continuum was created by means of PRAAT. The speaking rate was changed in equal-sized steps of 0.5 syll/sec from 2 syll/sec to 10 syll/sec. Sixteen German listeners participated in the experiment. They were asked to specify the rhythm of the presented stimuli by choosing one of the following three categories: (1) Strong syllable followed by weak syllable(s), (2) Weak syllable(s) followed by strong syllable, (3) uniform rhythm. The stimuli were played from individual desktop computers until the subjects made a decision by clicking on the corresponding icon on the screen. Reaction times were recorded together with the chosen rhythm categories.

3. RESULTS AND DISCUSSION

The results of the perception experiment are compatible with our assumption. At speaking rates above 8 syll/sec and below 3.5 syll/sec the sequences of strong-weak and weak-strong [b̥a] syllables show a significant increase of a uniform rhythm perception. Additionally, the rhythm perception is less robust, i.e. strong-weak sequences were more often perceived to be weak-strong sequences and vice versa. The reaction times increase significantly as well. In contrast, the sequences of physically identical [b̥a] syllables were clearly perceived to be rhythmically uniform above 8 syll/sec and below 3.5 syll/sec. However, for intermediate speaking rates, the identical [b̥a] syllables formed "subjective rhythms", i.e. the listeners clicked significantly more often on the weak-strong or strong-weak icons. The "subjective rhythm" perception was also linked with significantly longer reaction times. In conclusion, our study supports the existence of a rhythm window, and speakers comply with this window when producing emphatic and non-emphatic utterances. This compliance underlines the important role of speech rhythm in communicative functions (e.g, the guide function, emphasis) and in speech comprehension in general.

Speech rhythm: A mirage?

Francis Nolan and Hae-Sung Jeon

Department of Theoretical and Applied Linguistics, University of Cambridge
fjn1@cam.ac.uk

Keywords: speech rhythm, timing, pitch, prosodic typology

1. INTRODUCTION

It has generally been assumed that speech has rhythm, and our task is to describe and quantify it. This assumption arises naturally from perceptual impressions which differ between languages and can be rationalised in terms of distinct rhythmic templates (e.g. 'stress-timing' and 'syllable-timing'). The enterprise of quantifying rhythm was given a boost in the 1990s by the development of 'rhythm metrics' such as the Pairwise Variability Index (PVI) [6] and %V, ΔC [7]. It has also been assumed, reasonably enough, that to achieve quantification we need to focus on timing. This paper will question both assumptions, citing evidence of the interplay of timing with other dimensions (exemplified by pitch), and speculating that languages do not have rhythm, though the sensation of rhythm may arise epiphenomenally from speech in some – but not all – languages.

2. RHYTHM ISN'T (JUST) TIMING

Conceived within the broad framework of the 'stress-timed'~'syllable-timed' opposition, the PVI was first applied by Low [5] not only to timing but also to intensity and vowel dispersion. With hindsight, this was in the belief that the rhythmic percept emerges in the derived domain of prominence rather than any individual physical dimension. Low showed that the greater syllable-timing of Singapore English was indeed reflected in each of these dimensions. This early work did not, however, attempt to unravel their interaction. We will cite two studies which have.

In [1], Cumming explored the interaction of timing and pitch, from a low-level effect whereby pitch can affect perceived length, to a cross-linguistic difference in the prioritisation of the two dimensions in judgments of rhythmicality. The fact that language determines the nature of the interaction underlines the conclusion that rhythm is a construct. Ultimately, Cumming demonstrated a method for combining the two factors in a unified PVI.

Meanwhile Jeon [2] explored the interaction of pitch and timing in demarcating the Korean accentual phrase – so called despite the absence of clearly determinable prominences in Korean, but potentially the domain of rhythmic organisation most comparable to the foot. Findings will be discussed which show that timing and pitch may engage in a perceptual trading relation when the task is to detect an accentual phrase boundary, though the nature of the interaction depends on the specific question asked.

3. RHYTHM: AN EPIPHENOMENON?

Korean is not alone in lacking the unambiguous prominences which readily license a rhythmic percept; syllable prominences prove elusive in for instance in Tamil [4] and Mongolian [3]. Where prominences are evident, the analogy with clearly rhythmical phenomena such as music is persuasive, yet it is possible that this analogy is facilitated not by the presence of rhythm in the language, but by a chance conspiracy of non-rhythmic phonetic factors; that is, speech rhythm is epiphenomenal. We will develop this point by exploring how text aligns with music in diverse languages.

4. REFERENCES

[1] Cumming, R.E. 2010. The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica* 67(4), 219–242.

[2] Jeon, H-S. 2011. Prosodic phrasing in Seoul Korean: the role of pitch and timing cues. PhD Dissertation, University of Cambridge.

[3] Karlsson, A. 2005. Rhythm and intonation in Halh Mongolian. *Travaux de l'institute de linguistique de Lund 46*.

[4] Keane, E. 2006. Prominence in Tamil. *J. Int. Phonetic Association* 36, 1–20.

[5] Low, E.L., Grabe, E., Nolan, F. 2000. Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. *Lang. Speech* 43, 377–401.

[6] Nolan, F., Asu E.L 2009. The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66, 64-77.

[7] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.

Selective neuronal entrainment to the beat, spontaneously unfolded while listening to syncopated musical rhythms

Sylvie Nozaradan[1,2], Isabelle Peretz[2], André Mouraux[1]

[1] Institute of Neuroscience (Ions), Université catholique de Louvain (UCL), Belgium
[2] International Laboratory for Brain, Music and Sound Research, Université de Montréal, Canada
sylvie.nozaradan@uclouvain.be

Keywords: musical rhythm, neuronal entrainment, EEG, steady-state evoked potentials, beat perception, non-linear dynamics

1. INTRODUCTION

Feeling the beat is fundamental to the experience of music. This perception usually refers to the spontaneous ability to extract periodicities from rhythmic stimuli that are not strictly periodic in reality. This phenomenon is strikingly illustrated in syncopated rhythms, referred to in music as rhythmic patterns in which the perceived beat is not systematically concurrent to actual sounds (Large, 2008; Velasco and Large, 2011). Moreover, the beats are usually perceived within meters (e.g. a waltz, which is a three-beat metric structure).

How the brain spontaneously builds up internal representations of beat and meter while listening to music remains largely unknown. Recently, we provided direct evidence that listening to simple periodic beats elicits a neuronal entrainment which can be captured in the human electroencephalogram (EEG) as a beat-related steady-state evoked potential (SS-EPs), appearing at the exact frequency of the beat (Nozaradan et al., 2011). These results corroborate the resonance theory for beat perception (Large, 2008) which hypothesized that the perception of periodicities in music emerges from the entrainment of neuronal populations resonating at the frequency of the beat.

Here, using this novel approach, we investigated whether rhythmic patterns, whose temporal envelope contained multiple frequencies and which could be assimilated to syncopated rhythms, elicit multiple simultaneous SS-EPs at corresponding frequencies in the EEG spectrum. Most importantly, we aimed to examine (1) how a selective neuronal entrainment may emerge spontaneously at one frequency from these multiple SS-EPs, and (2) whether this selective neuronal entrainment may reflect beat-specific neuronal activities underlying the building of a beat representation in the human brain.

2. METHOD

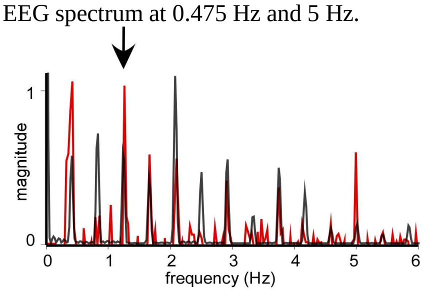
The EEG was recorded while nine normal participants listened to 5 distinct rhythmic patterns lasting 33 s. The patterns contained sequences of sounds alternating with silences such as to induce a perception of beat and meter based on the preferential grouping of 4 events at related metric levels. This was confirmed by a task performed at the end of the EEG recording, in which participants were asked to tap along the beat perceived along the rhythmic patterns.

For each pattern, the EEG frequency spectrum was computed (µV), as well as the spectrum of the rhythmic patterns envelope. This allowed assessing whether some frequencies would stand out specifically in the EEG compared to sound spectrum.

3. RESULTS

As predicted, the rhythmic stimuli elicited multiple SS-EPs at the frequencies corresponding to the temporal envelope of the rhythmic patterns. Most importantly, although the acoustic energy was not necessarily predominant at beat frequency, the amplitude of the SS-EPs elicited at the frequency of the perceived beat and meter was selectively enhanced as compared to the SS-EPs elicited at other frequencies.

Figure 1: EEG spectrum (red) and syncopated pattern envelope spectrum (black), normalized between 0 and 1. Frequency of the beat at 1.25 Hz is pointed by an arrow. Meter-related SS-EPs are also visible in the EEG spectrum at 0.475 Hz and 5 Hz.



4. DISCUSSION

As shown in Figure 1, the frequency corresponding to the beat was enhanced in the EEG spectrum, when compared to the envelope spectrum of the rhythms. Moreover, other frequencies were also enhanced, corresponding to metric structures. This indicates that SS-EPs elicited in the frequency range of musical tempo do not merely reflect the physical structure and magnitude of the sound envelope, but reflect a mechanism of neuronal selection of beat-relevant representation. Taken together, our findings suggest that studying the neural activity related to the perception of musical beats using SS-EPs constitutes a promising approach to gain insight on the general mechanisms of neuronal entrainment and its possible role in the formation of coherent representations of simultaneous streams of sensory information.

5. REFERENCES

[1] Large, E.W. 2008. Resonating to musical rhythm: Theory and experiment. In Simon Grondin (Ed.), The Psychology of Time, Emerald, West Yorkshire.

[2] Velasco, M.J., Large, E.W. 2011. Pulse detection in syncopating rhythms using neural oscillators. In Proceedings of the 12th annual conference of the international society for music information retrieval. In press.

[3] Nozaradan, S., Peretz, I., Missal, M., Mouraux, A. 2011. Tagging the neuronal entrainment to beat and meter. J. Neurosci. 31, 10234-10240.

Meter and rhyme ease processing time in poetry

Christian Obermeier[1], Martin von Koppenfels[2], Winfried Menninghaus[3], & Sonja A. Kotz[1]

[1] Research Group 'Subcortical Contributions to Comprehension', Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; [2] Institute of Comparative Literature, Ludwig Maximilians University Munich, Germany; [3] Institute of Comparative Literature, Free University of Berlin, Germany
obermeier@cbs.mpg.de

Keywords: meter, rhyme, oscillations, alpha, theta

1. INTRODUCTION

Rhetoric has long since claimed that rhythmical and metrical features of language substantially contribute to persuading, moving, and pleasing an audience. Similarly, literature largely relies on the presumed aesthetic and affective virtues of rhythm and meter (e.g. [1]). A potential explanation for these effects is provided by the cognitive fluency theory [2], which states that recurring patterns (i.e. meter) ease the cognitive processing of a stimulus and consequently enhance its aesthetic evaluation. Here, we explore the significance of meter and rhyme on the cognitive and affective processing of poetry by means of event-related brain potentials (ERPs). Participants listened to verses of German poems that varied in the aforementioned factors. We were particularly interested whether the ease of processing would be reflected in time-frequency patterns of the EEG.

2. METHOD

Participants
16 right-handed native speakers of German (7 female, all right-handed, mean age: 25)

Materials
A set of German verses was varied along two dimensions: meter (metered vs. non-metered) and rhyme (rhyming vs. non-rhyming). There were 60 verses per condition, resulting in a total set of 240 verses. All verses were spoken by a professional actor.

Procedure
Verses were presented in blocks of six per condition. After each verse participants judged the rhythmicity and overall aesthetic value of a verse.

Data Analysis
The behavioral data were each subjected to a repeated measures ANOVA. The EEG data were corrected for artifacts and filtered with a bandpass filter (0.5-30 Hz). The data were then transformed for the time-frequency analysis (TFA) using a wavelet-transformation. Separate cluster analyses were performed to determine potential Regions of Interest (ROIs) as well as potential frequency bands of interest for the effects of meter and rhyme. For these ROIs and bands (theta (4-8 Hz) and lower alpha (8-12 Hz), repeated measures ANOVAs were calculated with the factors meter (metered vs. non-metered) and rhyme (rhyming vs. non-rhyming).

3. RESULTS

The behavioral data show that both metered as well as rhyming verses are perceived as significantly more rhythmic and aesthetic than non-metered or non-rhyming verses (all paired *t*s $(15) > 4$; all ps $< .001$). The TFA revealed a main effect of rhyme in the theta-band, with non-rhyming verse eliciting higher activity than rhyming ones ($F(1,15) = 5.4$, $p < .05$. In the alpha-band there was a significant, right lateralized effect of meter with non-metered verses eliciting higher activity than metered ones (paired $t(15) = 2.44$, $p < .05$).

4. DISCUSSION

The results of the TFA show that both rhyme and meter independently affect the cognitive processing of poetry. Increased activations for non-metered and non-rhyming verses in the alpha and theta bands, respectively as well as the behavioral results suggest that meter and rhyme facilitate cognitive processing of poetry as proposed by the cognitive fluency theory.

5. REFERENCES

[1] Tsur, R. 2008: Towards a theory of cognitive poetics. Sussex Academic Press: Brighton and Portland.

[2] Reber, R., Schwarz, N., & Winkielman, P. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? Personality and Social Psychology Review 8, 364-382.

Integrating turn-taking behavior in a hierarchically organized coupled oscillator model

Michael O'Dell1[1]*, Tommi Nieminen*[2]*, Mietta Lennes*[3]

[1]School of Language, Translation and Literary Studies, University of Tampere, Finland
[2]School of Humanities, University of Eastern Finland, Finland
[3]Department of Modern Languages, University of Helsinki, Finland
michael.odell@uta.fi

Keywords: turn-taking, pauses, rhythm, coupled oscillators, synchronization

1. INTRODUCTION

A model of conversational turn-taking loosely based on coupled oscillators has been recently suggested by Wilson & Wilson [8]. Beňuš [1] tested several predictions of this model against a database of conversational American English. His results provided some support for the model, but the support was weak due to small (although significant) correlations, and a lack of predicted phase patterns.

As Wilson & Wilson pointed out, it is important to gather data on a variety of languages in addition to English. We recently began examining pausing and turn-taking behavior in a Finnish conversational database [7] using an analysis similar to that of Beňuš. Similar to Beňuš, our results were inconclusive.

2. METHOD

We use Bayesian nonparametric hazard function modeling (with the DPpackage in R [2]) to assess the possible effect of previous rhythms on the ongoing propensity to start speaking after pause, as well as the possible maintenance of silent synchronous turn cycles during pause. We also attempt to integrate turn-taking into our own coupled oscillator model of speech timing (cf. [3], [4], [5], [6]), which has hitherto lacked an explicit mechanism for dealing with pausing behavior.

3. RESULTS

Initial results suggest that Finnish speakers' behavior during pause is sensitive to previous rhythmicity, but not simply by adjusting an overall rate as implied by the Wilson & Wilson model. On the one hand there appears to be at most a single cycle of turn-taking lasting less than half a second at the beginning of pause. Previous speech rate does not appear to affect the rate of this cycle, but does correlate with overall propensity to start speaking, positively for the same speaker, negatively for the opposite speaker.

4. DISCUSSION

These results can be interpreted as indicating that pausing behavior represents an optimal trade-off between two processes, a short lived turn cycle as in the Wilson & Wilson model, and a slower general random process which dominates as pause continues. Oscillation could be a way to avoid simultaneous starts with short pauses. As pauses get longer, oscillation will be of less help because of deteriorating synchrony, and at the same time it will also be less needed.

5. REFERENCES

[1] Beňuš, Š. 2009. Are we 'in sync': Turn-taking in collaborative dialogues. – *Proceedings of the 10th INTERSPEECH*, 2167–2170.

[2] A. Jara, T. E. Hanson, F. A. Quintana, P. Müller, and G. L. Rosner. DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40(5), 2011

[3] O'Dell, M., Lennes, M., Nieminen, T. 2008. Hierarchical levels of rhythm in conversational speech. – *Speech Prosody 2008: Fourth Conference on Speech Prosody, Campinas, Brazil*, 355–358.

[4] O'Dell, M., Lennes, M., Werner, S., Nieminen, T. 2007. Looking for rhythms in conversational speech. – *Proceedings of the 16th International Congress of Phonetic Sciences*, 1201–1204.

[5] O'Dell, M., Nieminen, T. 1999. Coupled oscillator model of speech rhythm. – *Proceedings of the XIVth International Congress of Phonetic Sciences*, part 2, 1075–1078.

[6] O'Dell, M., Nieminen, T. 2009. Coupled oscillator model for speech timing: Overview and examples. –*Nordic Prosody: Proceedings of the Xth Conference, Helsinki 2008*. Peter Lang.

[7] O'Dell, M., Nieminen, T., Lennes, M. (in press). Modeling turn-taking rhythms with oscillators. – *Linguistica Uralica*.

[8] Wilson, M., Wilson, T. P. 2006. An oscillator model of the timing of turn-taking. – *Psychonomic Bulletin & Review*, 12: 957–968.

Pulse and the role of intrinsic frequency in temporal adaptation

Caroline Palmer[1,2], Pascale Lidji,[1,2] and Isabelle Peretz[2,3]

[1]McGill University; [2]BRAMS Lab; [3]University of Montreal; Canada
caroline.palmer@mcgill.ca

Keywords: phase adaptation; period adaptation; beat-deafness; pulse; intrinsic frequency

1. INTRODUCTION

Most humans, with or without training, exhibit fine synchronization as they clap to a song, hum with a tune, or sway to a beat. The fact that even untrained individuals respond with a strong sense of pulse is remarkable because these perceived periodicities arise even in response to music that is not periodic [1]. However, there are rare cases of individuals who cannot keep a beat [2]. What mechanisms account for this inability? We examine this question in a line of research that examines how nonmusicians and a "beat-deaf" individual entrain to a regular beat that introduces period and phase perturbations typical of those that occur in music.

2. METHOD

Following Large, Fink and Kelso [3], we presented phase and period perturbations in a metronomic stimulus to 29 musically untrained adult male participants who had to synchronize their taps with the changing beat. This control group was matched in age and gender and comparable in musical training to that of a previously reported beat-deaf individual, Mathieu [2], who performed the same task.

Participants were presented on each of six trials with a metronomic stimulus that first established a regular 500-ms beat period over 12-15 onsets, and then either increased or decreased its period or phase relative to the expected beat onset by 3, 8, or 15% of the baseline period. Each of these 12 perturbation types (period / phase X increase / decrease X 3, 8, 15%) were introduced over the course of each trial for 13-20 onsets, and were separated by a return to the baseline period of 500 ms. Each trial lasted a total of 306-331 onsets, and the synchronization measures (stimulus onset minus tap onset) were converted to relative phase values and averaged across trials.

3. RESULTS

The control group demonstrated significantly faster adaptation to phase perturbations than to period perturbations, as reported previously [3]. In contrast, the beat-deaf individual failed to adapt quickly to either perturbation type. Inspection of the response variability during the baseline period (in the absence of perturbations) indicated Mathieu's greater variability compared with the control group. A nonlinear dynamical systems model applied to the beat-deaf case indicated weaker fits of an intrinsic tapping frequency that, in the fits to control subjects, adapted in response to temporal perturbations.

4. DISCUSSION

Although most humans have a strong sense of pulse that underlies their ability to adapt to temporal perturbations in music, speech, and other auditory events, these findings suggest that some individuals lack that intrinsic periodicity. Temporal adaptation may rely on an entrainment of internal neural oscillations that capture the sensation of pulse present in most humans, regardless of specific musical training.

5. REFERENCES

[1] Large, E.W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science 26*, 1-37.

[2] Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia, 49*, 961-969.

[3] Large, E.W., Fink, P., & Kelso, S.J. (2002). Tracking simple and complex sequences. *Psychological Research 66*, 3-17.

Rhythmic jitter: Accounting for low nPVI in languages with quantity contrast

Viktória Papp

New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, New Zealand
viktoria.papp@canterbury.ac.nz

Keywords: rhythmic typology, phonotactics

1. INTRODUCTION

The occurrence of a phonemic short-long variation in a language would, in theory, enhance rhythmic variability, potentially pushing a language towards stress-timing. The extent to which this actually occurs, however, will be mediated by the frequency of occurrence of long vs. short segments. If long segments are fairly infrequent, or the durational difference between short and long segments is very small, it is expected that the effect would be attenuated.

This study is looking at three vowel harmony languages, and attempts to explain the rhythmic differences with the help of phonotactic and durational information about each language. The concept of rhythmic jitter is introduced and exemplified.

2. METHOD

In order to investigate the duration proportion of short and long segments in the two Finno-Ugric languages, a literature search was conducted. The duration proportion information was used to construct predicted nPVI values for the sentences in the Finnish, Hungarian and Turkish translation of Aesop's fable "The North Wind and the Sun." The predicted nPVI values were then compared to observed nPVI values based on the same passage. The difference between the predicted and observed nPVI values was modelled with jitter (noise) added to every duration value in the predicted PVI formula in 5% increments. As an example, if in the predicted PVI formula the length of one particular vowel was 1 unit, in the first round of jitter simulation its value was a value between 0.95-1.05 units, in the second between 0.9-1.1 units, etc. For each round of simulation, the predicted PVI was recalculated with the jitter-added values. The result of the jitter simulation is the jitter expressed in percentage that puts the predicted values the closest to the observed values. Finally, an analysis of written corpora in the three languages (~ 0.5M words / language) was carried out to quantify the proportion of short vs. long segments.

3. RESULTS

The literature search concluded that depending on the position in the word and elicitation methods, the duration of short and long segments was about 1:2.27-1:3 for Finnish vowels, about 1:1.5-1:1.8 for Hungarian vowels and about 1:2 for Finnish consonants. Regarding the frequency of short vs. long segments, for vowels Hungarian and Finnish are about on par with about three times as many short vowels as long ones and Turkish has overwhelmingly more short vowels than long ones (about 20 times as many). In the case of consonants the quantity contrasts between long vs. short segments are about 1:7 in Finnish, 1:14 in Hungarian, and 1:80 in Turkish.

As PVI is designed to capture unit-to-unit durational variability, it was predicted that the higher fluctuation in both duration and frequency of quantity distinctions disrupts isochrony and results in higher PVI values in Finnish than in Hungarian. Since Turkish has the least frequent incidence of long segments it was predicted that it has the most even unit-to-unit durational variation, that is, the lowest PVI, of the three languages. However, in the acoustic experiment, it was found that Turkish, a language without phonemic quantity contrast, shows a much larger interquantile range in both nVPVI% and nCPVI% than either Hungarian or Finnish, both of which exhibit prominent phonemic length contrast. Apart from a few outlying values, it is remarkable how compact both Hungarian and Finnish scatters are along both axes.

Following this, the predicted VPVI% values for the passage were calculated based on the durational data found in the literature and contrasted with the observed values. The jitter simulation showed that Hungarian introduces about 50-55% jitter, Finnish about 15-35% jitter and Turkish (depending on the short-long durational ratio) anywhere between 25-75% jitter.

4. DISCUSSION

If we define phonotactic prominence as the combined effects of durational and frequency ratio of short and long segments in a language then the results of this study show that the more prominent long segments are the less rhythmic jitter a language exhibits.

While this in itself is an interesting and somewhat counterintuitive fact, it is a phonotactic-phonetic fact as opposed to a claim about rhythmic typology, i.e. a claim about the existence and the location of relative isochrony. This puts pairwise metrics in a different light: from being a rhythmicity measuring tool they become a device to quantify the rhythmic jitter added to the predictable differences arising from the durational and frequency profile of a language.

After quantifying the large-scale differences at every level between the three languages the fact that the three languages pattern fairly closely comes as a surprise. Disregarding what the timing of these languages can be labeled as, it seems as if they introduce varying amounts of jitter just so that they can counterbalance the phonotactic differences.

Rhythmic typology of three languages with vowel harmony and quantitative contrast

Viktória Papp

New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, New Zealand
viktoria.papp@canterbury.ac.nz

Keywords: rhythmic typology, vowel harmony, quantity, pairwise metrics

1. INTRODUCTION

What makes the objective evaluation of the host of pairwise rhythmicity measures put forward from the 1990s is that they are rarely applied to the same language(s). This study therefore is applying the six most frequently used metrics to three languages: nCVPI%, nVPVI%, rCPVI%, %V, ΔC, VarCoΔC.

Another lacuna in this field concerns PVI-based typological categorization of languages with vowel harmony (VH) and quantitative contrast (QC). With the notable exceptions of Estonian [1] and Thai [2] these languages have been absent from the studies to date. Languages with quantitative contrasts are theoretically important to shaping this debate because the short-long distinction may introduce variability to segment durations, potentially shifting the language's rhythm typology towards the stress-timed end of the continuum. Vowel harmony languages, on the other hand, tend to exhibit very little or no vowel reduction, similarly to many syllable-timed languages. In the intersection of these two opposing forces stand languages that show both vowel harmony and quantity phenomena. The three languages in this study, Finnish, Turkish and Hungarian, exhibit both VH and QC and therefore it is predicted that they will pattern reasonably close to each other, with the difference correlated with their differences in QC.

2. METHOD

The nine language samples, three per language, were sourced either from the illustrations of the Handbook of the IPA [3] or were recorded in the sound booth of Rice University. The reading material was Aesop's fable "The north wind and the sun" translated into the three languages. The manual segmentation followed the method detailed in [2].

3. RESULTS

The three languages pattern together as the ones with the lowest ΔC after Estonian. When compared against measurements with regard to the proportion of the vocalic intervals (%V) in [4], the results are on par with Greek, Japanese, Welsh, Singapore English and Estonian among others. In the VarCoΔC metric the range of Finnish broadens, the Hungarian range narrows and the Turkish range maintains its size compared to ΔC. Turkish and Finnish are kept apart by this metric too, with Hungarian sprawling across both.

Along nVPVI% the three languages show great overlap with values that place them on par with mixed languages like Greek, Singapore English and Malay, as measured in [2]. rCPVI clearly divides the three languages. Hungarian and Finnish end up as the languages with the smallest pairwise variation after Estonian, while Turkish exhibits values similar to German, Welsh and Spanish. The nVPVI% x rCPVI% metric therefore classifies Finnish and Hungarian as somewhat non- prototypical syllable-timed languages (due to the medium as opposed to the expected low variation along VPVI), and Turkish as an even less prototypical syllable-timed language (due to small but appreciable shifts towards stress-timed values along both axes).

4. DISCUSSION

The slight differences between the three languages along vocalic metrics can be interpreted that vowel harmony and/or the lack of reduction keeps variation at a certain level, despite the phonemic length contrast in Finnish and Hungarian vs. Turkish. Estonian, Finnish and Hungarian, three languages exhibiting phonemic quantitative consonantal contrast, perhaps counter-intuitively, show very little pairwise variation in consonant durations. However, the short-long contrast, be it small, might be a very consistent fine phonetic distinction in these languages that these metrics may not be able to capture.

5. REFERENCES

[1] Asu, E.L. & Nolan, F., 2005. Estonian rhythm and the Pairwise Variability Index. In Proceedings, FONETIK 2005. Göteborg: Göteborg University, pp. 29-32.

[2] Grabe, E. & Low, E.L., 2002. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner, eds. Laboratory Phonology 7. Berlin: Mouton de Gruyter, pp. 515-546.

[3] International Phonetic Association, 1999. Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet, Cambridge: Cambridge University Press.

[4] Ramus, F., Nespor, M. & Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. Cognition, 72, 1-28.

Properties of professional drum accompaniment performances and their relation to listeners' groove

Johan Paulin, Kahl Hellmer, Guy Madison

Department of Psychology, Umeå University, SE-901 87 Umeå.
johan.paulin@psy.umu.se

Keywords: systematic variability, groove, onset detection, drumming

1. INTRODUCTION

Appreciation of music is strongly related to aspects of the performance. Here we ask to what extent the tendency to move to the music – groove – is related to performance properties of drum accompaniments, including dynamic and temporal variability.

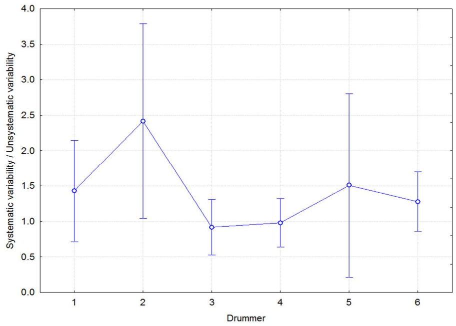
2. METHOD

Timing and dynamics of real drum performances was derived from studio recordings. The data consisted of 42 drum patterns culled from 26 performances by 7 drummers. Systematic variability was estimated in relation to the metrical structure. Groove was rated by 12 participants 20-34 years of age (M=26.08, SD=3.86).

3. RESULTS

Fig. 1 shows the ratio between systematic and unsystematic temporal variability, based on displacements from metric positions in the drum beat.

Figure 1: Systematic / unsystematic variability ratios for each drummer

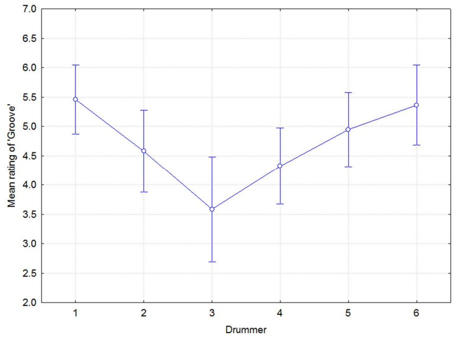


Listeners' ratings of the amount of groove in the performances are shown in Fig. 2. They were not significantly correlated with temporal parameters in the data; the correlation between groove and systematic variability was -0.16 and -0.14 with unsystematic variability (N=40). The only performance variable significantly correlated with groove ratings was the number of drum strokes per second (r=.38, N=39).

4. DISCUSSION

Here we show that simple drum accompaniments in popular rock music exhibit systematic variability patterns, in agreement with findings for other musical instruments and genres [2]. Although ratings of groove also differed significantly across both music examples and drummers, they were unrelated to the amount of systematic and unsystematic timing variability.

Figure 2: Groove ratings across drummers



This is consistent with a recent study that manipulated the amount of microtiming, empirically derived from real performances of jazz, funk, and samba music [1]. Similar conclusions obtain for correlations between ratings and rhythmic properties of full music examples across many genres. Rather, more groove was predicted by a salient beat and rhythmic richness [4], the latter in agreement with the present finding that the number of strokes per second was the only performance variable substantially correlated with groove. Groove has previously been attributed to microtiming and performance variability in general [e.g., 3]. This is based on the observation that performers and performances vary in their grooviness, at the same time as microtiming also varies among performers and performances. It would therefore seem a likely explanation that the one accounts for the other. We conclude that the source of groove must be sought elsewhere.

5. REFERENCES

[1] Davies, M., Madison, G., Silva, P., & Gouyon, F. (in preparation). The effect of microtiming deviations on the perception of groove in percussive music.

[2] Gabrielsson, A. (1999). Music performance. In D. Deutsch (Ed.), The psychology of music (pp. 501-602). San Diego: Academic Press.

[3] Iyer, V. S. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception*, *19*, 387–414.

[4] Madison, G., Gouyon, F., Ullén, F., & Hörnström, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *JEP:HPP, 37(5), 1578-1594.*

The influence of imposed metrical organization on temporal order judgments in isochronous sequences

Brandon Paul[1], *Per B. Sederberg*[2] *& Lawrence L. Feth*[1]

[1]Department of Speech and Hearing Science, The Ohio State University, USA
[2]Department of Psychology, The Ohio State University, USA
paul.674@osu.edu

Keywords: meter, strong and weak beats, temporal order, EEG

1. INTRODUCTION

Meter can be considered a higher-level attentional process whereby listeners organize temporally-spaced sound structures in hierarchies of strong and weak beats [1]. When acoustical features of these sounds do not vary, listeners can impose meter by grouping the sounds in different ways. Moreover, studies suggest that periodic sequences generate expectations in time that enhance perception [2]. Here, we investigate the possibility that listeners' temporal acuity is perceptually facilitated at times that are most strongly predicted, such that differences in temporal acuity arise between strong and weak beats using two types of metric organizations.

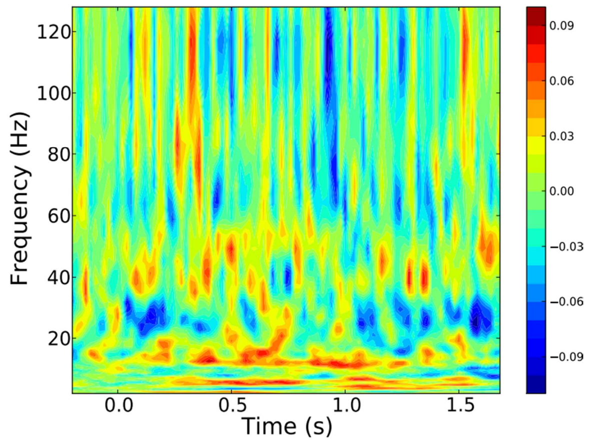
2. METHOD

While connected to an EEG recording system, 16 participants (9M, 7F) listened to isochronous sequences of stimuli, and are asked to impose either binary (strong-weak) or ternary (strong-weak-weak) meter. The subjects' task was to report the metric position on which deviant stimuli occur. Each stimulus consists of two closely-spaced wideband pulses of differing amplitude. A pulse pair that is reversed in temporal order is considered a deviant stimulus. A minimum of twelve stimuli were presented before the possibility of a deviant stimulus arose.

3. RESULTS

A main effect of beat type (strong or weak) was found when comparing proportion of correct identification of deviant stimulus metric positions ($F(2,19) = 8.34$, $p < .05$), but a post-hoc analysis only indicated differences on beat three in ternary meter versus all other beat types ($p < .05$). Preliminary neural analysis was conducted for one participant at a parietal electrode site for strong beats versus weak beats during stimulus sequences that precede the onset of a deviant stimulus. Figure 1 depicts spectral power differences between strong and weak beats, showing gamma band synchronization differences at a latency of 300ms. Observed differences may represent a metric imposition over identical stimuli.

Figure 1: Spectral power differences for strong and weak beats.



4. DISCUSSION

Although temporal acuity did not seem to be enhanced on strong beats versus weak beats, preliminary neural analysis shows that a mental representation of strong and weak beats in acoustically-identical stimulus sequences is observable in the EEG, providing evidence for neural correlates of imagined metric hierarchies. Results are implicated in furthering the understanding of cognitive mechanisms that help to organize complex temporally-dependent sound structures such as speech and music.

5. REFERENCES

[1] Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11,* 409-464

[2] Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time varying events. *Psychological Review, 106,* 119-159.

Timing and tempo in spontaneous self-repair

Leendert Plug and Paul Carter

Linguistics and Phonetics, University of Leeds, United Kingdom
l.plug@leeds.ac.uk

Keywords: phonetics, timing, error, self-repair, disfluency

1. INTRODUCTION

We report on a phonetic analysis of instances of self-repair taken from a corpus of spontaneous Dutch speech. We investigate instances of speech error repair in which a mispronunciation is corrected, and lexical repair, in which the 'reparandum' is an erroneous or infelicitous lexical choice. Our focus is on the relationship between repair timing — that is, when the repair is initiated relative to when the erroneous or infelicitous utterance has started — and repair tempo — that is, how the speaking rate of the repair compares to that of the reparandum.

Previous work suggests firstly that speech error repairs in which a mispronounced word is interrupted immediately after the erroneous segment has been produced have different pitch and intensity characteristics from repairs in which the mispronounced word is completed before repair is started [2]; and secondly that both speech error and lexical repair are commonly associated with an increase in speaking rate after the repair initiation [3]. The account of the prosodic differences between early-interruption repairs and late-interruption repairs [2] refers crucially to the degree of time pressure under which the repair is produced. Therefore, it would seem highly relevant to investigate their speaking rate characteristics, too.

2. METHOD

Our dataset consists of approximately 600 instances of self-repair sampled from the Spoken Dutch Corpus. We present separate analyses for speech error repairs and lexical repairs. Our approach in both cases is to model the speaking rate of the repair component on the basis of that of the reparandum, and assess whether adding additional predictor variables improves the fit of the model. Our focus, of course, is on the predictive value of measures of repair timing, although we also consider candidate predictors related to word frequency, phonological complexity and repair semantics, following [1].

We discuss several ways of quantifying repair timing and assess the extent to which they allow us to predict temporal details of the repair as a whole. For speech error repairs, for example, we can quantify the delay between error and repair onset with reference to the beginning of the word hosting the mispronunciation, or the word-internal point at which the mispronunciation first becomes overt. For both speech error and lexical repairs, we can take the repair onset to be the moment at which the utterance to be repaired is abandoned, or the start of the first correct lexical item.

3. RESULTS

Our analysis suggests that repair tempo is indeed constrained by repair timing, but whether significant effects are observed depends crucially on how repair timing is quantified. A proportional measure of reparandum completeness seems particularly effective as a predictor. Repairs are generally produced at a higher speaking rate than reparanda — but this speeding up is greater in instances with an interrupted reparandum. Repairs with interrupted reparanda are also likely to be produced with little delay between reparandum and repair.

4. DISCUSSION

Our results support the analysis in [2] that repairs produced early are under a different sort of time pressure from repairs produced later. This may reflect different monitoring procedures for pre- and post-articulatory monitoring.

5. REFERENCES

[1] Levelt, W.J.M., Cutler, A. 1983. Prosodic marking in speech repair. *Journal of Semantics* 2, 205-217.

[2] Nooteboom, S. 2010. Monitoring for speech errors has different functions in inner and overt speech. In: Everaert, M., Lentz, T., De Mulder, H., Nilsen, Ø. (eds), *The Linguistic Enterprise*. Amsterdam: John Benjamins, 213–233

[3] Plug, L. 2010. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of Phonetics* 39(3), 289–297.

Who's IN-TIME?

Kathinka Poismans

Research Centre KenVaK, university Zuyd, Heerlen, the Netherlands
kathinka.poismans@hszuyd.nl

Keywords: timing, autism, music therapy, rhythm, assessment instrument

1. INTRODUCTION

A newborn baby already has the ability to detect rhythmical patterns and to react to them rhythmically when interacting with its mother. An infant seems to have an inborn feeling for the rhythms of others and of the self which is a necessity when interacting with others [2]. Experience and research with autistic people, and the statements of autistic people themselves, indicate however that they have particular difficulty in identifying and entraining to the tempo and rhythm of another person and in timing their actions appropriately [1]. In music and in music therapy however they can respond to a beat, tempo and metre that is attuned by the music therapist to their needs and thus musical actions become more predictable and the possibility of "joining in" arises. An autistic person can, so to say, develop his feeling for his own and somebody else's rhythm. To study this process of gradually growing musical interaction, an observation instrument has been developed. It is called IN-TIME and it measures reciprocity in timing between music therapist and autistic client.

2. METHOD

The development of the instrument started with a qualitative study. Several video recordings of different music therapeutic settings, music therapists and clients were described, bringing into focus all aspects of timing within the musical action. Using techniques of Grounded Theory [3], 3 categories and 19 code rules were defined. Category 1 covers musical interaction without any coherence in timing. Category 2 covers musical interaction which displays rhythmical unison. Category 3 covers musical interaction which is rhythmically coherent but where the musical play of music therapist and child differ. Together they form IN-TIME, an instrument that can make a micro analysis of video recordings from music therapy sessions. 7 Raters were compared by reliability statistics. To check validity, the ability of the instrument to discriminate between autistic and non-autistic children was investigated. 10 Videos of autistic children making music with a music therapist were compared to videos of 10 neurotypical children and 10 mentally retarded children. The interventions were half standardized.

3. RESULTS

The inter-rater reliability test was executed by 5 raters who had undergone training and by 2 who had not. The agreement between trained raters was high (Cohen's Kappa between 0.57 and 0.93, mean 0.79).

The comparison between the three groups of children shows that all groups differ significantly in all categories (see table 1).

Table 1: Level of significance 0.05

ANOVA		Sig.
Duration in seconds cat.1	between groups	.000
Duration in seconds cat.2	between groups	.001
Duration in seconds cat.3	between groups	.000

4. DISCUSSION

IN-TIME was shown to be a reliable and valid instrument for measuring timing in the musical interaction between music therapist and autistic child. The analyses revealed different patterns of timing within the musical interaction for each group of children. Mentally retarded children showed mostly musical interaction of cat. 2. Neurotypical children showed few musical interactions of cat. 2 but very many in cat.3 which means that their timing is not only very good but also diversified. The autistic children show very little coherence in timing either of cat. 2 or cat. 3.

5. REFERENCES

[1] Poismans, K. (2009) Shared Time, timing in muziektherapie met autistische kinderen. *Wetenschappelijk Tijdschrift Autisme. Theorie en praktijk*, 1, p. 14-20.

[2] Stern, D. (2000) *The interpersonal world of the infant*. Basic Books.

[3] Strauss and Corbin, (1998) Basics of qualitative research. Techniques and procedures for developing Grounded Theory. London, Sage Publications.

Effects of stress and meter on duration perception in Spanish

Hugo Quené and Mariella Marquez Mendoza

Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands
h.quene@uu.nl

Keywords: metrical expectancy; stress; duration discriminability; Spanish; attention

1. INTRODUCTION

This study investigates whether perceptual discriminability of vowel duration is affected by the presence of lexical stress on that vowel, and by the listeners' metrical expectation of the vowel being strong or weak. A previous study using this paradigm for American English sentences [4] has reported error rates that confirmed the "attentional bounce hypothesis" [2]: discriminability was found to be higher for vowels in metrically predictable strong syllables. This effect of metrical (as opposed to rhythmical) expectations was previously reported by [2], but not found in a related study [3], while both studies had used lists of isolated words as stimuli. In addition, the American English discriminability study [4] also showed higher discriminability for trochaic patterns than for other metrical patterns, presumably because trochaic feet predominate in English.

2. METHOD

The present study reports on a parallel experiment on discriminability of vowel duration in meaningful Castilian Spanish sentences. Each stimulus sentence was 6 syllables long. The sentences varied in two important factors: metrical pattern (dactylus, trochee, iambus), and tempo (quick, slow). Each stimulus sentence contained a target vowel (/a,i,u/, in 3rd, 4th, or 5th syllable). Hence the target vowel was contained in a stressed (strong) or unstressed (weak) syllable, depending on the metrical pattern and serial position. Sentences were read by a male native speaker of Castilian Spanish. His average vowel durations were approximately equal between quick/strong and slow/weak vowels.

The target vowel was lengthened to 114%, 121%, 128%, 135% and 142% of its original duration, by means of PSOLA resynthesis in Praat [1].

Participants were 30 monolingual native speakers of Castilian Spanish, mostly students at the Technical Univ. Madrid. Their task was an AX forced-choice task, with the first member always having original durations. The order of stimulus pairs was randomized over participants. Both accuracy data (hits) and response times were collected.

3. RESULTS

The hit rates and response times were affected by the amount of lengthening and by overall tempo, which validates the experimental paradigm and procedure. Secondly, discriminability was affected by lexical stress, although the stress effect was considerably smaller in this Spanish study than in the previous American English study, and it did appear only weakly in the response times. This smaller effect of stress may be due to the non-reduction of unstressed vowels in Spanish. Thirdly, metrical patterns differed relatively little in their effects on duration discriminability, and performance was lowest for trochaic patterns. These smaller and qualitatively different effects of metrical patterns in the Spanish study may be explained by the absence of a predominant metrical pattern in Spanish.

4. DISCUSSION

These findings confirm the role of metrical expectations on listeners' sensitivity to phonetic details. These metrical expectations seem related to the occurrences of metrical patterns in a language, as well as to the phonetic realization of such metrical patterns in spoken language.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2010. Praat: Doing phonetics by computer [www.praat.org].

[2] Pitt, M.A., Samuel, A.G. 1990. The use of rhythm in attending to speech. *JEP:HPP* 16, 564–573.

[3] Quené, H. & Port, R.F. 2005. Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62, 1-13.

[4] Zheng, X., Pierrehumbert, J.B. 2010. The effects of prosodic prominence and serial position on duration perception. *JASA* 128, 851-859.

How listeners' respiration changes while listening to speech

Amélie Rochet-Capellan[1], Susanne Fuchs[1], Pascal Perrier[2]

1: Center for General Linguistics (ZAS) Berlin, Germany
2: GIPSA-lab/DPC Grenoble, France
ameliecapellan@free.fr

Keywords: breathing rhythm, listener-speaker interaction, convergence, loudness

1. INTRODUCTION

The beats of the heart and the respiratory cycles are probably the most fundamental biological rhythms in the human body. Our work focuses on the latter. Respiration does not only enable gas exchange for human biological functions, but also delivers the required air flow for speech. In dialogue situations, partners tend to synchronize their breathing cycles at turn takings [1]. So far it is unclear whether this synchronization occurs only in the vicinity of turn takings or whether it is a more general property of two dynamical systems getting coupled. In the latter perspective, listeners should couple their respiratory behavior with the speaker while listening. The main goal of our work is to investigate the potential coordination in breathing cycles between speakers and listeners. Our first hypothesis is that breathing patterns in a quiet environment are different from the ones produced while listening to speech. Our second hypothesis is that listeners adapt their respiratory patterns to changes in speakers' breath.

2. METHOD

We investigated how 26 female listeners adapted their breath to a female or a male speaker (called readers), reading short texts (fables) in loud vs. normal speech condition. The two readers differ in their respiratory behavior, in particular, in the way they realize the loud versus normal speech condition. Listeners' breathing patterns and acoustics were obtained with the same technique as for the readers. The listeners' tasks were the following: 1) to stand and breathe normally in quiet environment, 2) to read 5 short texts in their normal speaking style, 3) to listen to a reader via loudspeakers and briefly summarize the story afterwards. They either heard the male or the female speaker. Moreover, some speakers heard normal speech first and loud second and the others heard the reverse order.

3. RESULTS

Linear mixed model analyses were run on a variety of breathing parameters with condition (list_normal, list_loud, quiet), reader (male, female) and condition order (normal first or second) as fixed factors, and listeners and texts as random factors. In all these analyses, quiet breathing turned out to be significantly different from listening to speech, supporting hypothesis 1. Quiet breathing profiles look generally bell-shaped whereas it looks more skewed in the listening to speech conditions. Furthermore, an effect of loud versus normal speech was found, (hypothesis 2) which differed among the subjects hearing the male or female reader. Listeners hearing the female reader showed larger amplitudes of in/exhalation in loud versus normal speech. Listeners hearing the male reader showed the reverse pattern. This result may be interpreted with respect to the breathing kinematics of the male. When speaking loud, he showed an extreme deep inhalation and long duration of a breathing cycle. Female listeners were possibly not able to synchronize with such a behavior and therefore realized two breathing cycles within one cycle of the speaking male. A more detailed synchronization analysis is currently carried out.

4. DISCUSSION

Listeners' adaptation to speaker's breathing pattern has been found in our experiment. Thus, the respiratory synchronization between interlocutors seems not limited to turn taking. We are far away from knowing the explicit reasons for the studied adaptation, but several explanations might be possible: The influence of speakers on listeners may be a process of imitation and not primarily speech related. Converging behavior may also correspond to higher cognitive process between interlocutors which is bound to successful communication [2].

5. REFERENCES

[1] McFarland, D.H. 2001. Respiratory markers of conversational interaction. *JSLHR* 44, 128-143.

[2] Stephens, G.J., Silbert, L.J. & Hasson, U. 2010. Speaker-listener neural coupling underlies successful communication. *PNAS* 107(32), 14425–14430.

Movement timing is affected by differences in interval-specifying sounds

Matthew W. M. Rodger and Cathy M. Craig

School of Psychology, Queen's University Belfast, Belfast, UK
m.rodger@qub.ac.uk

Keywords: movement, timing, continuation paradigm, sound

1. INTRODUCTION

Guiding the timing of movements can be extrinsic (using sensory information to control the closure of temporal gaps) or intrinsic (generating temporal information neurally) [2]. Evidence of this distinction has been found in a sensorimotor synchronization task in which participants were more accurate when timing movements between alternate target zones to discrete beats (intrinsic guidance during intervals), but more consistent when moving to looping rising tones (extrinsic information about closing time gaps) [3].

Differences in the reproduction of time intervals specified by different sound types were investigated using a continuation paradigm, to provide insight into how the mind processes time when specified either intrinsically and extrinsically.

2. METHOD

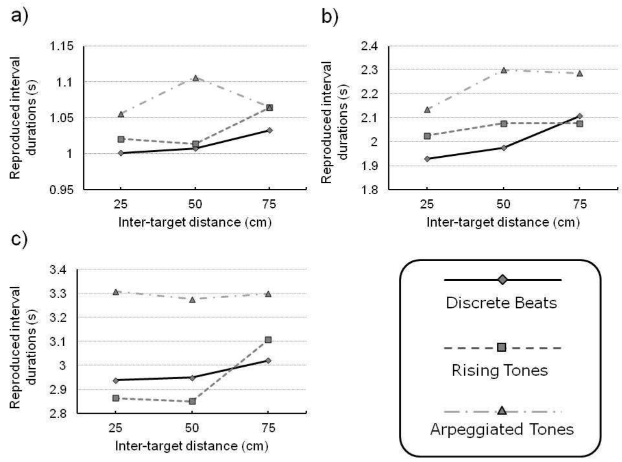
Three different auditory guides (discrete beats, rising tones, arpeggiated tones) for three interval durations (1s, 2s, 3s) were created. Discrete beats were 50ms tones (C6) at the onset of each interval. Rising tones were pitch-modulated tones increasing from C4 to C6. Arpeggiated tones were consecutive discrete tones rising chromatically from C4 to C6.

Thirty-eight adults (Females = 23; age range = 18-50 years) moved alternately between different targets (25cm, 50cm and 75 cm apart) in time with the presented sounds. After 16 repetitions, the sounds stopped, at which point participants had been instructed to continue moving between targets at the same pace, until told to stop after 16 continuation repetitions. The order of trials (3 sound types x 3 durations x 3 inter-target distances) was randomised.

3. RESULTS

Preliminary analysis (n=10) of reproduced time intervals are shown in Figure 1. For all interval durations, there was an effect of sound type on mean reproduced movement times (1s: $F_{(2,18)} = 26.1$, $p < 0.001$; 2s: $F_{(2,18)} = 16.3$, $p < 0.001$: 3s: $F_{(2,18)} = 6.4$, $p = 0.009$). Post-hoc tests revealed that these effects were due to the arpeggiated tone intervals being reproduced as longer than the other sounds. Although larger movement distances tended to produce longer durations, this effect was not yet significant.

Figure 1: Mean reproduced interval durations for each sound type across different inter-target distances for specified interval durations of: a) 1s; b) 2s; c) 3s.



4. DISCUSSION

These initial results show that reproduction of time intervals is affected by the extrinsic sensory patterns through which timing is specified. One possible interpretation of longer intervals produced in the arpeggiated tones condition is that the multiple discrete tones are processed as individual events, each of which has its own finite neural decay time [1]. This may mean that subjective time was expanded due to the higher density of auditory events within the same objective time interval.

A full analysis of all participants, including kinematic analysis of inter-target movements, will be presented.

5. REFERENCES

[1] Karmarkar, U., Buonomano, D. 2007. Timing in the absence of clocks: Encoding time in neural network States. *Neuron* 53, 427-438.

[2] Lee, D. 1998. Guiding movement by coupling taus. *Ecological Psychology* 10, 221-250.

[3] Rodger, M., Craig, C. 2011. Timing movements to interval durations specified by discrete or continuous sounds. *Experimental Brain Research* 214, 393-402.

New approaches to prosodic analysis: Galician song

Rosalía Rodríguez-Vázquez

Department of English, French and German, University of Vigo, Spain
rosalia@uvigo.es

Keywords: text-setting, prosody, phonology, Galician, Spanish

1. INTRODUCTION

This paper deals with the connections between the prosody of a language and the text-setting rules of vocal music in that language. The underlying hypothesis states that the rhythmic structure of a language will be reflected in the musical setting of texts arisen spontaneously in that language.

Research on linguistic rhythm relies on a three-fold classification, according to which languages may be syllable-timed, stress-timed or mora-timed [5; 1]. The notion of rhythmic classes has been questioned by those who favour the distribution of languages along a rhythmic continuum [3; 2]. Whatever the view taken, there is a general tendency to regard Germanic languages as stress-timed, while Romance languages are commonly placed at the syllable-timed end of the continuum. English is thus classified as a stress-timed language, while Spanish is viewed as totally syllable-timed by some [4] and as less syllable-timed than French by others [6].

The present paper aims at determining the mutual relationships between the prosodic characteristics and the text-setting rules of Galician, a Romance language spoken in the northwest of the Iberian Peninsula which stands between Portuguese and Spanish, showing a mixed type of rhythm.

2. METHOD

A theoretical and empirical analysis of a corpus of 50 Galician folk songs is undertaken in order to determine the extent to which Galician and Spanish differ with regard to the structure and function of stress. By looking at how a text is set to music we can shed some light on the connection between vowel reduction, the realisation of stress and the structure of stress-groups in Galician, comparing the results to those obtained for Spanish [7].

3. RESULTS

Despite the fact that Galician and Spanish are Romance languages spoken in the same territory, text-setting in Galician and in Spanish show diverging characteristics which are arguably derived from the prosody of the languages. Galician and Spanish show no reduction of prominent vowels. However, in Galician there is reduction and sometimes even deletion of unstressed vowels, a feature shared with European Portuguese but not with Spanish.

In Spanish song, the key issue is the preservation of syllabic rhythm even if this entails ‘musically conditioned stress shift’. In Galician song, this type of stress shift is disfavoured, as it would result in the misplacement of reduced unstressed syllables.

4. DISCUSSION

The initial hypothesis is confirmed, as the prosodic characteristics of the languages analysed here get reflected in the rules of text-setting. The definition of a musically conditioned stress shift in Spanish responds to syllable-related constraints, whereas it is linked to stress in Galician. Stemming from that, the agreement between linguistic stress and musical beat is prevalent in Galician (as in English), not in Spanish.

As this paper shows, vocal music constitutes an ideal locus for the analysis of rhythm in language.

5. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*, Edinburgh: Edinburgh University Press.

[2] Bertinetto, P. M. 1989. Reflections on the Dichotomy ‘Stress’ vs. ‘Syllable-timing’. *Revue de Phonétique Appliquée*, 91/93, 99-130.

[3] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalysed. *Journal of Phonetics*, 11, 51-62.

[4] Navarro Tomás, T. 1918. *Manual de pronunciación española*. Madrid: Gredos.

[5] Pike, K. L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

[6] Pointon, G. E. 1980. Is Spanish really syllable-timed? *Journal of Phonetics*, 8, 293-304

[7] Rodríguez-Vázquez R. 2010. *The Rhythm of Speech, Verse and Music: A New Theory*. Bern: Peter Lang.

Testing rhythmic and timing patterns of Italian

Antonio Romano[1]*, Paolo Mairano*[2]

[1] LFSAG, Università di Torino, Italy [2] GIPSA-Lab, Université Grenoble 3, France
[1]antonio.romano@unito.it

Keywords: speech rhythm, syllable and foot compression, Italian.

1. INTRODUCTION

The role of syllable compression in Italian has been investigated by [3] whose results indicate that it is a base temporal unit, both for compression and position effects. The shortening of the nucleus and of the onset in correspondence of an increase in syllable complexity suggests that the controlled temporal unit extends all over the syllable. Other data give further support to the hypothesis that this unit is the entire V-to-V temporal interval.

These results are partially in contrast with previous findings by e.g. [4] and [6] who studied patterns of compression on different speech samples.

This paper is aimed at testing on data of Italian (including sentences like the ones in [3, 4, 6]) the speech rhythm model proposed by [5] and [1] which predicts temporal patterns as the result of the coupling of two oscillators. Within this framework, the duration of the inter-stress interval (*ISI*) is a function of the number of syllables and of the coupling strength of two clocks.

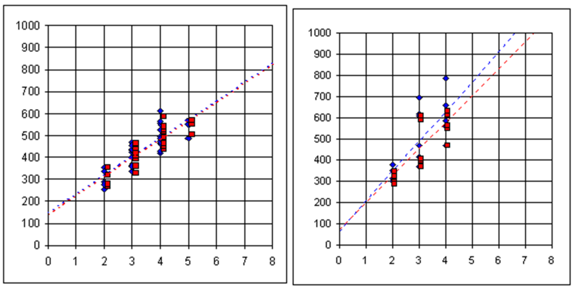
2. METHOD

We designed a small corpus similar to the ones analysed in these studies and we tested the model on 18 sentences read by 5 native Italian speakers (and on similar French and English sentences recorded and analysed in the same way). The sentences were segmented and labeled with Praat. Their rhythmic properties were assessed in terms of *Delta*s and *PVI*s and a regression analysis was carried out on *ISI*s.

3. RESULTS

Partial results for the first two speakers we have already analysed are summarised in Figure 1 together with the regression lines giving estimates of a roughly linear growth. For data on the left plot (speaker 1) the *a* parameter of the coupled-oscillator model has quite high values (145÷148) whereas the *b* parameter varies around 85. Data on the right plot (speaker 2) have *a* spanning from 69 to 78 and a higher *b* (127÷141). The coupling strength *r* shows mean values between 1.14 and 1.44 for speaker 1 and between 0.48 and 0.70 for speaker 2, accounting for a stable dominance of the syllabic oscillator: these variations are greater than the ones induced by changes in speech rate.

Figure 1: The growth of *ISI* in the data of two Italian speakers. Duration (ms, y-axis) vs.Number of syllables (*n*, x-axis).



4. DISCUSSION

In compliance with what is usually claimed for Italian, the partial results of the regression analysis we carried out for data of two speakers indicate a roughly linear growth in both cases. Likewise, the *Delta*s and the *PVI*s indicate low values of vocalic and consonantal variability for these samples. As a general result, which could be confirmed by the forthcoming analyses, Italian data are taken apart from data of English and, in some way, from data of French, even though both rhythm metrics and coupling parameters show in some cases only a weak discriminating power.

5. REFERENCES

[1] Barbosa, P.A. 2006. *Incursões em torno do ritmo da fala*. Campinas, Pontes.

[2] Bertinetto, P.M. 1983. Ancora sull'italiano come lingua ad isocronia sillabica. In: *Scritti linguistici in onore di G.B. Pellegrini*, II, Pisa, Pacini, 1073-1082.

[3] Farnetani, E., Kori, Sh. 1986. Effects of Syllable and Word Structure on Segmental Durations in Spoken Italian. *Speech Communication* 5, 17-34.

[4] Marotta, G. 1985. *Modelli e misure ritmiche: la durata vocalica in italiano*. Bologna, Zanichelli.

[5] O'Dell, M., Nieminen, T. 1999. Coupled oscillator model of speech rhythm. *Proc. of the XIVth ICPhS*, San Francisco, 1075-1078.

[6] Vayra, M., Avesani, C., Fowler, C. 1984. Patterns of temporal compression in spoken italian. *Proc. of the Xth ICPhS*, Utrecht (1983), 2, 541-546.

Rhythmic development in Spanish-English bilinguals

Elaine Schmidt, Brechtje Post

University of Cambridge
aeis2@cam.ac.uk

Keywords: rhythm, 1st language acquisition, bilingualism

1. INTRODUCTION

While many attempts have been made to quantify rhythm perceptions with the help of rhythm metrics, the acquisition of the features that contribute to rhythm remains an under-researched area, especially when looking at bilinguals with typologically different prosodic backgrounds. Previous research in monolinguals has shown a complex picture [1], including common developmental paths across languages with different prosodic typologies, as well as language-specific patterns emerging as early as age 2, indicating that phonological and prosodic features interact [2].

2. METHOD

In this paper we report two studies which investigated the rhythmic development in Spanish-English balanced bilingual 2-, 4- and 6-year olds living in the UK and Spain. In these studies, we tested rhythm development in semi-structured elicitation tasks in which the children described everyday actions. We asked whether the children follow the same developmental paths as monolinguals, or whether target-like acquisition of rhythm in one will facilitate rhythm acquisition in the other language, as has been found for segmental features [3]. The question arises whether in bilingual children the systemic properties that contribute to rhythm develop in parallel or independently of each other. We analysed metrical structures, vowel reduction, pre-boundary lengthening and accentuation together with a set of rhythm metrics, found to be discriminative for child speech in previous studies [2] (Varco-V, ΔC, %V, rPVI-C; cf. [4]).

3. RESULTS

Initial results show that bilinguals like monolinguals master the less complex syllable-timed rhythm of Spanish earlier than stress-timing. Indeed, their vocalic measurements in Spanish are comparable to those of monolinguals. However, even though they are generally slower in English with regards to vocalic metrics, they are still faster in their development than monolinguals. At the age of 4, %V is already on target in bilinguals while English monolinguals only reach adult-like results at 6. However, the development of vocalic metrics also shows a clear influence of the ambient language, with bilinguals living in Spain producing off-target vocalic durations still at the age of 6.

Interestingly, bilinguals outperform monolinguals of both language groups in the acquisition of consonants. While both Spanish and English monolinguals still show too much variability in their consonant lengths (ΔC and rPVI-C), bilinguals are approaching the adult target at the age of 4 and are completely on target by 6 in both languages.

4. DISCUSSION

These initial findings suggest that rhythm development is multisystemic; and in bilinguals the systemic properties of the two languages interact, possibly because of a greater variety of structures which monolinguals do not have access to. This suggests, perhaps unsurprisingly, that rhythm development crucially depends on the language systems that are present in the input. However, because monolinguals in the structurally more complex language also develop more slowly than bilinguals, it appears to be the greater structural variety in the input which serves to speed up acquisition, rather than structural complexity itself.

5. REFERENCES

[1] Grabe, Esther et al. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings of the International Congress of Phonetic Sciences*, (pp. 1201-1204). San Francisco.

[2] Payne, E. et al. (2012). Measuring child rhythm. *Language and Speech*. *55*. 203-229

[3] Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism*, *5*(1), 19-38.

[4] Post, Brechtje et al. (2010) "A multisystemic model of rhythm development: Phonological and prosodic factors". BAAP 2010. London, UK

Early electrophysiological markers of formal and temporal prediction

Michael Schwartze[1], Nicolas Farrugia[1,2], and Sonja A. Kotz[1]

[1]Max Planck Institute for Human Cognitive and Brain Sciences
Research Group "Neurocognition of Rhythm in Communication", Leipzig, Germany
[2]WSFiZ in Warsaw, Department of Cognitive Psychology, Warsaw, Poland
schwartze@cbs.mpg.de

Keywords: temporal structure, regularity, attention, EEG, deviance processing

1. INTRODUCTION

There is rhythm if we can predict on the basis of what is perceived [1]. Predictions pertain to the formal (e.g., color or pitch) and/or temporal structure of events in the environment, i.e. their "what" and "when" aspects. Crucially, predictable temporal structure may optimize processing of formal structure. It affects later stages of attention-dependent auditory deviance processing (Schwartze et al., 2011). Here, we discuss the impact of temporal predictability on two early components of the event-related potential of the Electroencephalogram (EEG), namely P50 and N100.

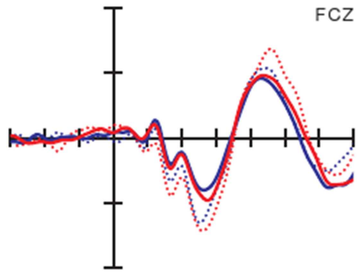
2. METHOD

Temporally predictable (isochronous 600 ms inter-stimulus-interval) and non-predictable (random 200-1000 ms inter-stimulus-interval) auditory oddball sequences were presented to 24 right-handed university students (12 female, mean 24.4, SD 2.8 years) in two experimental sessions. The sequences consisted of 512 standard (600 Hz) and 128 deviant (660 Hz) equidurational (300 ms) sinusoidal tones. In a pre-attentive session, participants focused on a silent video, while they focused on the sequences in a subsequent session. During both sessions, EEG and electrooculography were recorded.

3. RESULTS

The results indicate, that the amplitude of the P50 is sensitive to formal and temporal predictability, and to temporal regularity in particular. A high degree of predictability is associated with smaller P50 amplitude. P50 and N100 amplitudes are correlated in both sessions. Formal and temporal structure interact in the N100 obtained in the attention-dependent session.

A)



B)

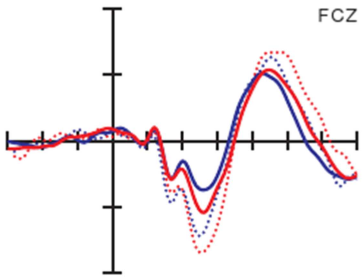


Fig 1: Group-averaged ERPs obtained in the pre-attentive (A) and in the attention-dependent (B) session reflecting responses to isochronous (solid blue) and random (solid red) standard, as well as isochronous (dotted blue) and random (dotted red) deviant tones.

4. DISCUSSION

The P50 is an early automatic marker of formal and temporal predictability, while the N100 may be influenced by attention. The results show that the P50 amplitude is smallest for attended isochronous standards and largest for attended random deviants. This specific pattern may reflect the relative effort to allocate attentional resources to the formal and temporal structure of the environment.

5. REFERENCES

[1] Fraisse, P. 1982. Rhythm and tempo. In: Deutsch, D. (Ed.).The psychology of music. Academic Press, New York, 149-180.

[2] Schwartze, M., Rothermich, K., Schmidt-Kassow, M., Kotz, S.A. 2011. Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology* 87, 146-151.

Slave to the rhythm: The roles of auditory processing in speech production

Sophie K. Scott, Carolyn McGettigan and Zarinah Agnew

Institute of Cognitive Neuroscience, University College London
sophie.scott@ucl.ac.uk

1. INTRODUCTION

Speech production is a complex motor task involving motor cortex, sub-cortical and cerebellar areas. Strikingly, posterior-medial auditory fields are strongly recruited during speech production, even if the speech is silently mouthed. In contrast to more anterior areas associated with the comprehension of speech, this auditory involvement in speech production has been linked to sensory-motor interactions – where the sensation could be auditory and/or somatosensory – in the control of speech output (Wise et al, 2001, Warren Wise and Warren 2005). The posterior-medial auditory areas seem to perform cross modal sensory-motor interactions during speech perception and production, and thus may form a central link between these two processes. This region is not driven solely by speech: it has been implicated in the rehearsal of music information (Hickok et al, 2003), and a direct comparison of speech production with non speech sound production, such as blowing a raspberry or whistling reveals very similar activation in this region (Chang et al, 2009).

In this presentation the role of posterior auditory areas in the control of timing in speech production will be addressed, using examples of speech production under conditions of auditory delay, as well as examples of the control of speech production in highly structured rhythmic speech. I will argue that auditory (and somatosensory) representations are important in the act of speaking and that these posterior medial auditory areas form part of an important sensory-motor system that is central to the production (and perception) of rhythm in speech. This may relate to its apparent non-specificity to linguistic information, since we can produce rhythms with any kind of vocalization and other kinds of sound output (e.g. in music).

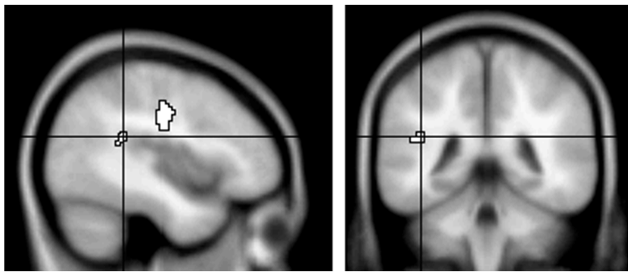
2. METHOD

Positron Emission Tomography (PET) and functional magnetic resonance imaging (fMRI) were used in several experiments to identify the neural systems involved in speech production tasks.

3. RESULTS

Figure 1 shows activation in left medial posterior auditory cortex during silent and normal speech production (Wise et al, 2001). Activation in this area increases during the production of nursery rhymes (Blank et al, 2002) and when speaking under conditions of delay (Takaso, Eisner et al, 2010).

Figure 1: The activation of an auditory area, posterior and medial to primary auditory cortex, which is activated by speaking, mouthing, and voicing (Wise et al, 2001).



4. DISCUSSION

The results suggest that posterior medial auditory fields are intimately involved in speech production. I shall show how the responses in auditory areas may differ from those seen in motor speech areas and discuss the candidate roles these different systems may play in speech perception and production, especially with reference to the use of speech in conversation.

5. REFERENCES

[1] Blank SC, Scott SK, Murphy K, Warburton E, Wise RJ. (2002) Speech production: Wernicke, Broca and beyond. Brain; 125(8):1829-38.

[2] Chang SE, Kenney MK, Loucks TM, Poletto CJ, Ludlow CL. (2009) Common neural substrates support speech and non-speech vocal tract gestures. Neuroimage. 47(1):314-25

[3] Hickok G, Buchsbaum B, Humphries C, Muftuler T. (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci, 15(5):673-82.

[4] Takaso H, Eisner F, Wise RJS and Scott SK (2010) The effect of delayed auditory feedback on activity in the temporal lobe while speaking: A PET study. Journal of Speech Hearing and Language Research. 53(2): 226-36.

[5] Warren JE, Wise RJ, Warren JD. (2005) Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci, 28(12):636-43.

[6] Wise, RJS, Scott, S. K., Blank, S. C, Mummery, CJ, Warburton, E (2001) Identifying separate neural sub-systems within 'Wernicke's area', Brain, 124, 83-95.

How well can speakers produce temporally-regular taps and syllables?

Stefanie Shattuck-Hufnagel[1] *and Lauryn Zipse*[2]

[1]Speech Communication Group, RLE, MIT, Boston, Massachusetts, USA, and [2]MGH Institute of Health Professions, Department of Communication Sciences and Disorders, Boston, Massachusetts, USA
sshuf@mit.edu

Keywords: rhythm, speech, aphasia, tapping, syllable repetition

1. INTRODUCTION

There is growing interest in the role of rhythm in speech processing [3,5] and in speech therapy [1,4,6], but little is known about the distribution of various rhythmic capacities, Our battery of rhythm tasks tests such skills in typical young adults, older people with aphasia, and age-matched controls. Here we report results for 2 simple tasks: repeated production of finger-taps and a syllable.

2. METHOD

Nine people with aphasia (PWA; mean = 61.3, SD = 9.9), 9 age-matched controls (mean = 63.2, SD = 18.0), and 8 younger controls (range = 18-22) completed 3 trials each of tapping and of repeating the syllable *puh* at a steady comfortable rate. For each speaker, the mean and SD of the inter-tap interval (ITI) and of the inter-syllable interval (ISI) were found. To quantify individual variability in production rates, the coefficient of variation (CV; SD/mean) was computed per speaker.

3. RESULTS

Variability was considerable within each group for ITI, ISI, and variability of these two measures. The within-group CV was 0.418 for PWA, 0.305 for older controls, and 0.334 for younger controls. Tapping and syllable production rates were slowest for the PWA, but were not significantly different from either of the two control groups with this sample size (Fig. 1). Generally, such isochronous repetition was faster for tapping than for syllable production (5/9 PWA, 7/9 older controls, and 8/9 younger controls). Intra-subject variability was relatively low and similar across groups (Fig. 2).

Figure 1: Mean Inter-tap and Inter-syllable Intervals.

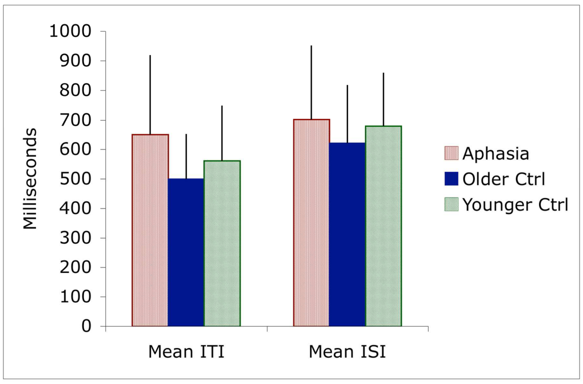
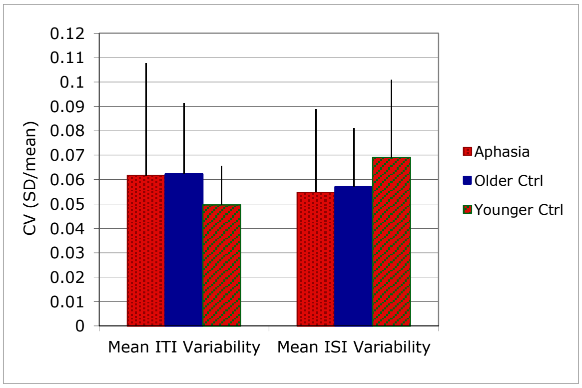


Figure 2: Mean Coefficient of Variation for ITI and ISI.



4. DISCUSSION

Temporal variability is higher in various clinical groups [2], but here the control groups are as variable as PWA (Fig. 2). In this first study of spontaneous isochronous syllable production, the rate for this task was slower than the tapping rate (although this difference was less systematic in PWA). This difference might be because speech is a more complex motor act, or because tapping provides simpler sensori-motor feedback than speaking CV syllables does.

5. REFERENCES

[1] Boucher, V. et al. 2001. Variable efficacy of rhythm and tone in melody-based interventions: implications for the assumption of a right-hemisphere facilitation in non-fluent aphasia. *Aphasiology* 22, 131-149.

[2] Corriveau, K. H., Goswami, U. 2009. Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex* 45, 119-130.

[3] Cummins, F., Port, R. F. 1996. Rhythmic commonalities between hand gestures and speech. *Proc. 18th Annual Conf. of Cognitive Science Society.*

[4] Glover, H., et al. 1996. Effect of instruction to sing on stuttering frequency at normal and fast rates. *Percepual & Motor Skills* 83, 511-522.

[5] Goldstein et al. 2012. Coupled oscillator model of speech timing and syllable structure, http://sail.usc.edu/~lgoldste/ArtPhon/Papers/Week%2012/ LG_final.pdf

[6] Konczak, J., et al. 1997. Control of repetitive lip and finger movements in Parkinson's Disease: influence of external timing signals and simultaneous execution on motor performance. *Movement Disorders* 12, 665-676.

Biases toward metrical (ir)regularity in speech production

Jason A. Shaw

MARCS Institute, University of Western Sydney, Australia
J.Shaw@uws.edu.au

Keywords: word stress, lexical priming, word naming, OCP

1. INTRODUCTION

Studies of stress placement in nonce words have revealed systematic biases towards metrical regularity, both in the production of sentences [e.g., 1] and isolated words in priming paradigms [e.g., 2]. In this study, a priming paradigm was used to investigate the effect of repeating metrical patterns on real word production. Results provide evidence both for the familiar sub-lexical bias toward metrical regularity as well as a new effect—avoidance of consecutive identical lexical stress patterns.

2. METHOD

Participants (24 native English speakers) were asked to read English words of the form CVCVC (e.g., *cabin, parrot, gazelle, saloon*), from a computer screen as quickly as possible. The words, consisting of **iambic primes** (25 medium frequency iambs such as *cadet*, *parade*, *ravine*, *sedan*), **trochaic primes** (25 medium frequency trochees such as *caret, palace, ribbon, ballad*), and **iambic targets** (10 low frequency iambs such as *gazette*, *lapel*, *meringue*, *pipette*), were displayed one at a time following the hextuple structure in table 1 until all target iambs were produced twice in each priming condition.

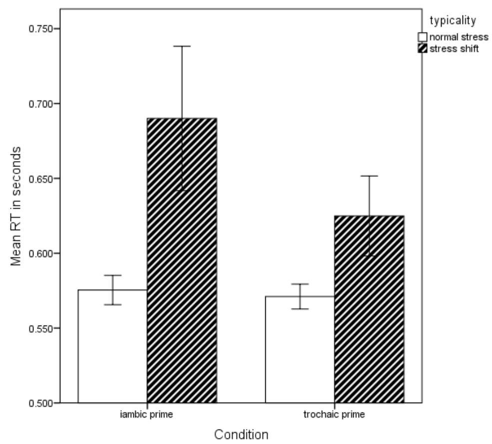
Table 1: Structure of the stimuli presentation

hextuple structure	prime condition	
	incongruent	congruent
(trial 1) random	random	random
(trial 2) prime	trochee	iamb
(trial 3) prime	trochee	iamb
(trial 4) prime	trochee	iamb
(trial 5) target	**iamb**	**iamb**
(trial 6) random	random	random

3. RESULTS

Target iambs were produced accurately in a practice block by all participants. In the experimental conditions, stress shifted to the first syllable of target iambs in 18% (180/972) of the trials. These cases of stress shift occurred more often in the iamb condition (104/472, 22%) than in the trochee condition (76/500, 15%). Fig 1 shows the reaction time (RT) for target iambs produced with normal stress (white bar) and with stress shifted to the first syllable (striped bar) in both conditions. When stress was shifted, RT was faster in the trochee condition than in the iamb condition.

Figure 2: Reaction times for normal and stress shifted targets by condition (iambic prime vs. trochaic prime)



4. DISCUSSION

Repeated metrical structure facilitated production only when words were produced without the lexically specified stress pattern (c.f., stress priming in Italian [3]). In these cases, RT was faster when stress shifted in the direction of the primes. This likely reflects the same sub-lexical mechanism involved in nonce word stress placement [1, 2]. In addition, the results yielded a bias for metrical irregularity. There was a greater tendency to shift stress when target iambs were preceded by other iambs than when preceded by trochees. This bias can be seen as well in the reaction times for responses to prime words, which were reliably faster and less variable on trial 2 (mean = 467ms.; SD = 27ms.) than on trial 3 (mean = 474ms.; SD = 40ms.). The bias against repeated patterns of lexical stress is a new result and one that may be related to the tendency for grammatical classes to differ in metrical structure.

5. REFERENCES

[1] Kelly, M. & Bock, J. 1988. Stress in time. *JEP: HPP* **14**(3): 389-403.

[2] Colombo, L. & Zevin, J. 2009. Stress priming in Reading and the Selective Modulation of Lexical and Sub-Lexical Pathways. *PLoS One* **4**(9).

[3] Sulpizio, S., Job, R., Burani, C. 2011. Priming lexical stress in reading Italian aloud. *Language and Cognitive Processes*: 1-13.

Timing in talking: What is it used for, and how is it controlled?

Alice Turk and Stefanie Shattuck-Hufnagel

Linguistics & English Language; U. of Edinburgh, and Research Laboratory of Electronics, MIT
turk@ling.ed.ac.uk

Keywords: timing, duration, speech production, motor control

1. INTRODUCTION

Timing is an integral part of every aspect of speech production: of individual movements of the rib cage, tongue, jaw, lips, velum and laryngeal structures, of their coordinated muscular activity, and of the speech sounds they produce. An understanding of speech production therefore requires an understanding of timing: 1) what it is used for, and 2) how it is controlled. In the first part of this paper, we review our current understanding of what speakers use timing for, and how this understanding was acquired. We propose that one of the main uses of speech timing is to make utterances easier to recognize: it is used to signal individual speech sounds (e.g. *did* vs. *dad*) [1], and also to signal, and compensate for, the relative predictability of syllables and words due to their context and frequency of use (e.g. [2], [3], [4], [5]). We propose that this recognition goal is balanced against other goals, such as the need to speak quickly, or in rhythm, to yield surface sound durations in speech. We highlight the important role of prosodic structure for speech timing: Prosodic structure serves as the interface between language and speech [6],[7],[8], and controls acoustic saliency so that it compensates for relative (un)predictability [3],[4],[5]. In the second half, we focus on two different views of how timing is controlled, i.e. with and without a domain-general timekeeping mechanism. Theories such as DIVA [9], based on VITE [10], and many Optimal Control Theory approaches (e.g. [11]) assume a general timekeeping mechanism, whereas Articulatory Phonology/Task Dynamics [12-15] suggest mechanisms for achieving surface timing patterns without a domain-general timekeeper. We present timing phenomena that occur in both speech and non-speech, showing how they can be explained within each type of framework. We finish by presenting evidence that may be difficult to explain without a domain-general timekeeping mechanism. This evidence includes greater timing variability for longer duration intervals compared to shorter duration intervals (e.g phrase-final segments vs. phrase-medial segments, [16]), patterns of differential timing variability for movement onsets vs. target attainment [17], and data suggesting a constraint on maximum syllable durations for phonemically short vowels in Northern Finnish [18].

2. REFERENCES

[1] Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei in English. *JASA 32*(6), 693-703.

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

[3] Aylett, M. 2000.Ph.D.thesis, University of Edinburgh.

[4] Aylett, M., & Turk, A. (2004). *Lang. and Speech, 47*, 31-56.

[5] Turk, A. (2010). Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Journal of Laboratory Phonology, 1*, 227-262.

[6] Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic Prosody II* (pp. 111-140). Trondheim: TAPIR.

[7] Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193-247.

[8] Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics, 101*, 112-156.

[9] Guenther, F. H. (1995). Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural-Network Model of Speech Production. *Psychological Review, 102*(3), 594-621.

[10] Bullock, D., & Grossberg, S. (1988). Neural Dynamics of Planned Arm Movements - Emergent Invariants and Speed Accuracy Properties during Trajectory Formation. *Psychological Review, 95*(1), 49-90.

[11] Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience, 5*(11), 1226-1235.

[12] Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic: structure. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). New York: Academic Press.

[13] Saltzman, E. L., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382.

[14] Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*(2), 149-180.

[15] Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Paper presented at Speech Prosody 2008, Campinas, Brazil.

[16] Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *JPhon, 26*(2), 173-199.

[17] Perkell, J. S., & Matthies, M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/ - within-subject and cross-subject variability. *JASA 91*(5), 2911-2925.

[18] Nakai, S., Turk, A., Suomi, K., Granlund, S.C., Ylitalo, R. & Kunnari, S. (under review). Quantity and constraints on the temporal implementation of phrasal prosody in Northern Finnish.

Timing in models of speech motor control

Alice Turk and Stefanie Shattuck-Hufnagel

Linguistics & English Language; U. of Edinburgh, and Research Laboratory of Electronics, MIT
turk@ling.ed.ac.uk

1. INTRODUCTION

Articulatory Phonology/Task Dynamics currently provides the most comprehensive account of speech timing phenomena. In this talk, we provide an in-depth review of speech timing in this model, and contrast it with the current instantiation of DIVA/Vite (which could be extended in the Optimal Control Theory framework). Articulatory Phonology/Task Dynamics and DIVA control timing in very different ways: In Articulatory Phonology/Task Dynamics, surface timing properties emerge from intrinsic characteristics as well as a set of timing adjustment mechanisms. In DIVA (and in most Optimal Feedback Control theories), desired surface movement times are specified through the use of a domain-general timekeeper. Although Articulatory Phonology/Task Dynamics is currently unsurpassed in its ability to account for most speech timing phenomena, general timekeeper approaches have desirable properties which suggest that they too deserve consideration and development. These properties include the ability to specify desired surface movement durations as part of a movement plan, as well as the ability to independently prioritize the timing of different movement components relative to each other or to another event (e.g. target attainment vs. movement onset).

2. REFERENCES

[1] Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic: structure. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). New York: Academic Press.

[2] Saltzman, E. L., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382.

[3] Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*(2), 149-180.

[4] Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Paper presented at Speech Prosody 2008, Campinas, Brazil.

[5] Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *JPhon, 26*(2), 173-199.

[6] Guenther, F. H. 1995. Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural-Network Model of Speech Production. *Psychological Review, 102*(3), 594-621.

[7] Bullock, D., & Grossberg, S. (1988). Neural Dynamics of Planned Arm Movements - Emergent Invariants and Speed Accuracy Properties during Trajectory Formation. *Psychological Review, 95*(1), 49-90.

[8] Todorov, E., & Jordan, M. I. 2002. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience, 5*(11), 1226-1235.

[9] Simko, J., & Cummins, F. (2010). Embodied Task Dynamics. *Psychological Review, 117*(4), 1229-1246.

Investigating the role of adaptation and anticipatory mechanisms on sensorimotor synchronization using ADAM

M.C. (Marieke) van der Steen [1], Merle T. Fairhurst [1] and Peter E. Keller [1,2]

[1] Max Planck Research Group "Music Cognition and Action", Max Planck Institute for Human Cognitive and Brain Sciences, Germany
[2] MARCS Institute, University of Western Sydney, Australia
steen@cbs.mpg.de

Keywords: sensorimotor synchronization, temporal adaptation, anticipation, modeling

1. INTRODUCTION

Sensorimotor synchronization (SMS) is the temporal coordination of self-generated motor rhythms with external rhythmical events. Temporal adaptation and anticipation mechanisms have been proposed to be involved in SMS [1]. We investigated the role of these mechanisms with help of an Adaptation and Anticipation model (ADAM) that synchronized with recordings of human tapping sequences.

2. METHOD

ADAM combines an established formal model of adaptation (phase and period correction) [2] with an anticipation process inspired by the notion of internal models [3]. Participants first learned to tap a tempo changing sequence resembling accelerando and ritardando in performed music. The recorded taps functioned as input for ADAM. By changing the settings of ADAM, three different models were created (Table 1). Variables of interest are SMS accuracy (mean absolute asynchrony) and SMS variability (standard deviation of asynchronies).

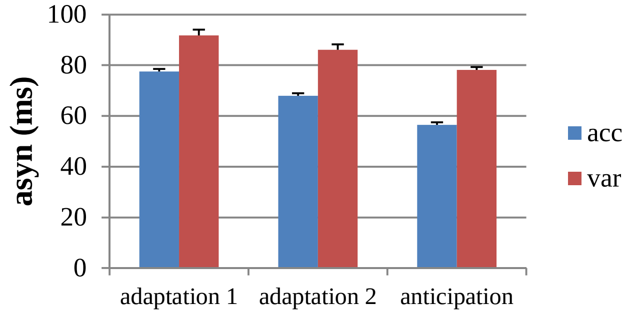
Table 1: Active mechanisms in the three different models

MODEL	MECHANISM		
	phase correction	period correction	prediction
adaptation 1	x	-	-
adaptation 2	x	x	-
anticipation	x	-	x

3. RESULTS

Preliminary results of 1 participant show a beneficial effect of period correction and prediction on SMS accuracy ($F_{2;28}$ = 74.58, $p<0.001$) and variability ($F_{2;28}$ = 24.14, $p<0.001$) (Figure 1).

Figure 1: Preliminary results: SMS accuracy and variability



4. DISCUSSION

The results suggest that adapting to and predicting event timing in a tempo changing sequence both facilitate synchronization. In a follow up experiment the effect of adaptation and anticipatory mechanisms on SMS in a bidirectional set up will be investigated.

5. REFERENCES

[1] Keller, P.E. 2008. Joint action in music performance. In: Morganti F, Carassa A, Riva G (eds) Enacting intersubjectivity: a cognitive and social perspective to the study of interactions. IOS Press, Amsterdam, pp 205-221.

[2] Wolpert, D.M., Kawato, M. 1998. Multiple paired forward and inverse models for motor control. *Neural Networks* 11, 1317-1329.

[3] Repp, B.H., Keller, P.E. 2008. Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science* 27, 423-456.

Idiosyncrasies in local articulation rate trajectories in Czech

Jan Volín [a)], *Lenka Weingartová* [b)]

[a)] Department of Anglophone Studies, Metropolitan University Prague, [b)] Institute of Phonetics, Charles University in Prague, Czech Republic
volin@mup.cz

Keywords: duration, articulation rate, cluster analysis, semantics

1. INTRODUCTION

Temporal structure of speech has to reflect both the properties of the language that have developed over the course of its history, and the individual patterns of the speaker's production habits.

In speaker recognition tasks, separation of the two sources of surface speech forms is the key to success. However, current methods of speaker identification or recognition make almost no use of larger scale temporal features [1, but cf. 2]. One of the few attempts to integrate temporal characteristics into speaker recognition systems is the SAUSI (Semi-Automatic Speaker Identification) system developed by H. Hollien [3]. It utilizes information about temporal structure such as the number and length of pauses, the ratio of speech to pauses, the ratio of speech time to total time, syllable rate, etc.

Our goal is to find such methods of describing Czech temporal patterns that would be less 'global' than those used by Hollien and more linguistically interpretable than those employed in current technological speech processing.

2. METHOD

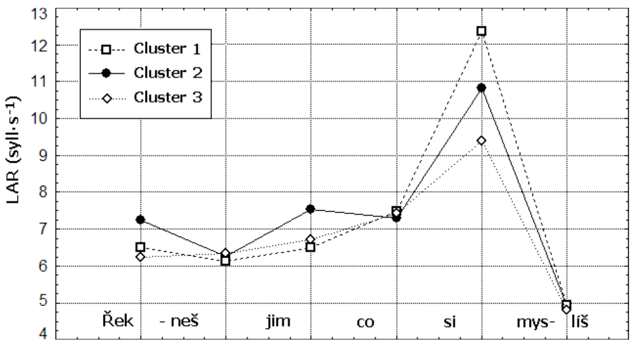
Short scripted dialogues (5 turns in each) were rehearsed and read out in a recording studio by a total of 24 speakers. On separate days the speakers read different roles in each dialogue. The target sentences were placed in the third turn of each dialogue and were constructed to differ from each other in only two respects: statement vs. question, and verb in singular vs. verb in plural. We analyzed two of such quadruplets of sentences, i.e., 192 items (24 × 4 × 2).

From automatically segmented and manually corrected material, LAR measure (Local Articulation Rate, see [4]) was extracted to obtain a temporal contour. The extracted contours were normalized by the group mean and subjected to cluster analysis (k-means method with constant distances of initial centroids) to see if there were any patterns of accelerations and decelerations that would group speakers into distinct sets.

3. RESULTS

The analysis identified moments in the temporal contours at which the speakers differed most. E.g., the utterance in Fig. 1 (*Řekneš jim, co si myslíš*) shows that final lengthening on the autosemantic *myslíš* (i.e., *think*) is a poor differentiator between the speakers while the treatment of synsemantic *si* (i.e., *self*) can split the group of speakers into types.

Figure 1: A local articulation rate contour of the utterance *Řekneš jim, co si myslíš,* i.e., *You'll tell them what you think.*



Other analyzed sets produced similar outcomes. Interestingly, no significant differences between statements and questions were found: speakers tend to be consistent in producing certain specific temporal patterns for both grammatical categories.

4. DISCUSSION

Cluster analysis shows what is not clearly detectable from raw data or compressed parameters like means and standard deviations. The results are encouragingly consistent. It seems possible for Czech to identify words, morphemes and their contexts that are temporally less rigid and reveal speakers' rhythmic preferences.

5. REFERENCES

[1] Hermansky, H. 1998. Should recognizers have ears? *Speech Communication* 25, 3–27.

[2] Hermansky, H., Sharma, S. 1998. TRAPS – Classifiers of temporal patterns. In: Proceedings of ICSLP'98, Sydney, Australia.

[3] Hollien, H. 2002. *Forensic Voice Identification.* London: Academic Press.

[4] Volín, J. 2009. Metric warping in Czech newsreading. In: R. Vích (ed.) *Speech Processing* 19, 52–55.

Effects of imitative training techniques on L2 production and perception

Ewa Wanat, Rachel Smith and Tamara Rathcke

University of Glasgow, UK
ewa.wanat@hotmail.co.uk

Keywords: L2 learning, L2 production, L2 perception, synchronous speech, imitation

1. INTRODUCTION

Imitation of a native speaker has been claimed to improve both production and comprehension [1] of a second language (L2). A recent suggestion [2] is that imitation in synchrony with a target speaker is particularly beneficial because it simultaneously engages the production and perception systems and provides immediate feedback on performance. We investigated how two types of imitation affected L2 learning of phonological contrasts in production and perception. Polish learners of English were tested on production and perception of English segmental contrasts before and after being exposed to a native speaker's production.

2. METHOD

15 Polish learners of English, resident in Glasgow for between 1 to 6 years, were exposed to a set of 18 sentences spoken by a speaker of SSBE containing instances of two English features that are difficult for Polish learners: the contrast between tense /i/ and lax /ɪ/, and voicing of word-final, utterance-final consonants [3]. Each sentence was put into a loop of 8 repetitions. The exposure task differed for the 3 groups of participants (5 per group): one group listened to the loops (listen-only or LO group), the second repeated each sentence after the target speaker (listen-and-repeat or LR group), and the third repeated synchronously with the target speaker (repetitive synchronous imitation or RSI group).

Before and after exposure, subjects did perception tests (a modified AXB task and an intelligibility-in-noise test) and read a set of control sentences containing novel words with the key features. These tasks allowed us to test which method was most successful in improving participants' perception accuracy and L2 pronunciation respectively.

3. RESULTS

The perception results showed an effect only for the LO group, who significantly improved in the post-exposure perception task. This contrasts with Adank *et al.*'s [1] finding that vocal imitation improves language comprehension. The post-exposure production test showed that both groups involved in production during exposure were closer to the target in the post-test than the LO group. Further, as far as the vowel duration is concerned, the RSI group accommodated better to the target speaker's production during the exposure task than the LR group, whereas the LR group were better than the RSI group at generalising vowel duration to the new (control) sentences. A significant result was also found for the exposure data, namely that the RSI group diverged less from the target speaker than the LR group on normalised F1 and F2 as well as duration values. The RSI group generalised the formant changes better in the post-test, i.e. showed better vowel quality learning.

4. DISCUSSION

Taken together, these results suggest that a combination of both perception training through exposing the subjects to a native speaker's voice and production training through synchronising with the target speaker could be a successful method of teaching L2 perception and production.

5. REFERENCES

[1] Adank, P., Hagoort, P., Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science,* 21(12), 1903-1909.

[2] Harrer, G. (1997) "Effectivare spr°akundervisning med ny metod," <http://www.diu.se/nr1-97/nr1-97.asp?artikel=rsi > (accessed on 14/03/2011).

[3] Koźbiał, S. (2011) Phonological error mapping: an English – Polish contrastive study. <http://atp.uclan.ac.uk/buddypress/diffusion/volume-4-issue-1/phonological-error-mapping-an-english-polish-contrastive-study > (accessed 08/08/2011).

Where is the rhythm in speech?

Laurence White

School of Psychology, Plymouth University, UK
laurence.white@plymouth.ac.uk

Keywords: prosodic timing, rhythm class, segmentation

1. INTRODUCTION

The isochrony hypothesis has long been disproved [2], but still informs conceptions of speech rhythm in two important ways. Firstly, the notion that languages can be divided into a small number of discrete "rhythm classes" remains pervasive, despite the weakness of the phonetic evidence [3]. Secondly, it is widely assumed that the temporal structure of speech is rhythmical, with successive groups of sounds coordinated in time according to higher metrical structure [1,5]. This paper reports speech perception studies which challenge these conceptions and support a functional and fundamentally non-rhythmical approach to the modelling of prosodic timing processes.

2. STUDY A: DISCRIMINATION WITHIN AND BETWEEN RHYTHM CLASSES

Utilising speech modified to focus on temporal properties, studies have shown that infants and adults can discriminate languages between, but not within, rhythm classes [4,6]. We present a series of experiments showing that adult listeners, English and Italian, can discriminate within rhythm class, and indeed within a single language, where there are linguistic differences in syllable rate and in localised lengthening effects. Such results argue against the notion of categorical rhythm classes and in favour of gradient, functionally-motivated distinctions.

3. STUDY B: LOCALISED LENGTHENING CUES TO WORD SEGMENTATION

Studies of artificial language learning have shown that speech timing strongly mediates listeners' ability to segment words based on statistical regularities. Specifically, segmental lengthening is interpreted across languages as a cue to the end of a perceptual unit, a process that is held to be universal and essentially non-linguistic [7]. We report a series of experiments that demonstrate the perceptual importance of linguistic localisation of durational cues, with word onset cued by consonantal lengthening and word offset by vowel lengthening, challenging the idea that durational cues are diffuse, epiphenomenal or indeed non-linguistic.

4. A FUNCTIONAL APPROACH TO PROSODIC SPEECH TIMING

These and other findings regarding the production and interpretation of timing effects are evaluated against the functional criteria that communicative durational variation should be consistent and specifically localised, and furthermore perceivable and interpretable by the listener [8]. In the resulting functional taxonomy, prosodic structure is held to influence speech timing directly only at the heads and edges of prosodic domains, through large, consistent lengthening effects. As each such effect has a characteristic locus within its domain, speech timing cues are seen to be less ambiguous for the listener than would be the case if the duration of segments were consistently determined with reference to higher metrical structure.

5. REFERENCES

[1] Barbosa, P.A. (2007). From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Communication*, *49*, 725-742.

[2] Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.

[3] Loukina, A., Kochanski, G., Rosner, B., Shih, C., Keane, E. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, *129*, 3258-3270.

[4] Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766.

[5] Port, R.F. (2003). Meter and speech. *Journal of Phonetics*, *31*, 599-611.

[6] Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *Journal of the Acoustical Society of America*, *105*, 512-521.

[7] Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, *126*, 367-376.

[8] Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*, 329-336.

Optimization modeling of speech timing

Andreas Windmann, Juraj Šimko, Britta Wrede, Petra Wagner

Faculty for Linguistics and Literary Studies, Bielefeld University, Germany
andreas.windmann@uni-bielefeld.de

Keywords: speech timing, computational modeling, optimization

1. INTRODUCTION

We present a novel paradigm for the computational modeling of speech timing. Our approach is based on the assumption that much of the variation in speech can be explained as emerging from trade-offs between minimizing effort and maximizing perceptual clarity, as proposed by H&H theory [2]. This is implemented as an optimization procedure, minimizing a composite cost function whose components relate to durations of various prosodic constituents. The model extends a recent computational implementation of this idea, which shows that various temporal coordination phenomena at the level of articulatory gestures can be accounted for on these grounds [4].

2. MODELING AND RESULTS

Our model operates on sequences of syllables representing prosodic phrases. The temporal organization of these sequences emerges as the result of an optimization procedure resolving trade-offs between production and perception constraints. The production-related component cost function *D* increases logarithmically with syllable durations, providing a crude measure of production effort. Simultaneously, a rational function *P* of syllable durations provides an impetus to lengthen syllables in a non-linear fashion, taking into account syllabic complexity. This function is meant to model perceptual constraints, approximating the inverse of the probability of recognizing a syllable *s* of a given duration [4]. Timing of higher-level prosodic units is evaluated as a cost related to difference functions *S* and *F* of syllable and inter-stress interval durations, respectively, implementing the assumption that these levels compete for governing the rhythmic organization of speech [3]. Moreover, a cost *T* on the duration of a phrase provides a control mechanism for speech tempo. These components of the cost function are combined as a weighted sum *C*. This facilitates to model phenomena such as stress and final lengthening by locally varying premiums imposed on individual components. Stressed syllables, for example, are modeled by temporarily increasing the perception cost weight π.

Formally, the model definition can be written as

$$C = \sum_{s} \alpha_s D_s + \alpha P \sum_{s} \pi_s P_s + \alpha_T T + \alpha_S S + \alpha_F F$$

where the $\alpha's$ represent weighting factors of the individual component cost functions.

In a preliminary experiment, we have constructed artificial “corpora”, representing the typical distributions of syllable types in different languages [1]. Given appropriate weights for the higher-level prosodic cost functions motivated by traditional rhythmic characterizations of the individual languages, the model reproduces language-specific timing phenomena, namely regression coefficients for inter-stress interval duration as a function of the number of syllables [3] and foot-level shortening. Importantly, the inclusion of syllabic complexity in the model ensures that the output does not exhibit isochrony at any prosodic level but reflects realistic durational variability.

3. DISCUSSION

Our results show that the proposed approach holds promising prospects for the modeling of speech timing. In particular, it provides a strong platform for *explaining* timing phenomena, applying general principles that have been shown to hold in other speech domains. Our approach thus represents a step towards the development of a unified account of the relationship between production-perception trade-offs and the variability encountered in speech.

4. REFERENCES

[1] Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

[2] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J. & Marchal, A. (eds.), *Speech Production and Speech Modeling*. Dordrecht, Kluwer, 403-439.

[3] O‘Dell, M., Nieminen, T. Coupled oscillator model of speech rhythm. *Proc. XIVth ICPhS*, 1075-1078.

[4] Simko, J., Cummins, F. 2010. Sequencing and optimization within an embodied task dynamic model. *Cognitive Science* 35(3), 527-562.

"…And I feel good!" Subjective, neural and embodied correlates of the pleasure of groove

Maria. A. G. Witek,[1] Eric Clarke[1], Mikkel Wallentin[3], Mads Hansen[3], Morten Kringelbach[2] & Peter Vuust[3]

[1]Faculty of Music, University of Oxford, UK. [2]Dept. of Psychiatry, University of Oxford, UK.
[3]Center for Functional Integrative Neuroscience, University of Aarhus, Denmark.
maria.witek@wadh.ox.ac.uk

Keywords: groove, syncopation, body-movement, pleasure.

1. INTRODUCTION

What is it about music that makes us want to move? And why does it feel so good? Few contexts of musical enjoyment make the pleasurable effect of music more obvious than in a dance club. A growing body of research demonstrates that music, traditionally conceived of as a cultural artefact, activates areas of the brain known to be involved in the regulation of rewards that have clear biological values, such as food and sex. Music's ability to stimulate expectations is believed to explain this effect. However, the role of body-movement and dance in pleasurable responses to groove-based music, such as funk, hip hop and electronic dance music, has been ignored in such theories. In this study, we investigated whether varying degrees of syncopation in groove affects the desire to move and feelings of pleasure using a number of methods.

2. METHOD

Stimuli: 15 synthesised funk drum-breaks, 2-bar loops, 120 bpm, 16 seconds, with varying degrees of syncopation (0-85) according to a polyphonic index of syncopation [1].

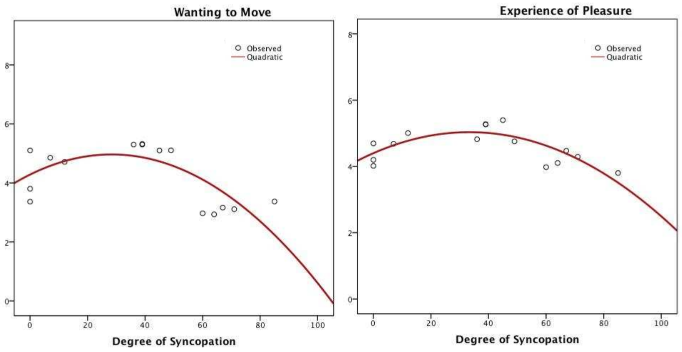
fMRI Study: Neural responses were recorded in 26 participants using functional Magnetic Resonance Imaging, while listening to and rating the drum-breaks according to how much they made them want to move and how much pleasure they experienced.

Motion-Capture Study: After scanning, participants were asked to moved freely to the drum-breaks in a motion-capture lab while movement-acceleration and -velocity was recorded using Nintendo Wii and WiiDataCapture (one sensor in the right hand and one on the back).

3. RESULTS

We found that a quadratic curve could explain 51.7% and 62.1% of the variation in movement and pleasure ratings respectively (Figure 1). Using SPM, this inverted U-shaped curve could be reflected in patterns of increased Blood Oxygenated Level Dependent (BOLD) responses in the auditory cortex, while a negative linear relationship was found in the basal ganglia. Furthermore, early explorations of the motion-capture data suggest similar trends with regards to degree of syncopation and kinetic force of movements.

Figure 1: Ratings of wanting to move and experience of pleasure fit with quadratic curve.



4. DISCUSSION

This triangulation of subjective, neural and embodied responses suggests that an intermediate degree of syncopation in groove facilitates body-movement and pleasure. We thus provide unique insights into the rewarding and movement-eliciting properties of music, which have been used to explain the cultural ubiquity and biological origin of music. As few can resist the urge to tap their feet, bop their heads or get up and dance when they listen to groove, such insights are a timely addition to theories of music-induced pleasure and entrainment.

5. REFERENCES

[1] Witek, M. A. G., Clarke, E. F., Kringelbach, M. L., Wallentin, M., Hansen, M. & Vuust, P. 2011. The effects of polyphonic context on syncopation in music. *Conference on the Neurosciences and Music IV: Learning and Memory,* University of Edinburgh, 9-12 June 2011.

Evidence for coordination between overlap onsets and inter-stress intervals

Marcin Włodarczak, Juraj Šimko, Petra Wagner

Faculty of Linguistics and Literary Studies, Bielefeld University, Germany
juraj.simko@uni-bielefeld.de

Keywords: dialogue, temporal entrainment, overlapped speech

1. INTRODUCTION

Rhythmic factors have been suggested to play a role in managing the timing of speaker changes in dialogue. For example, [1] proposed that speakers entrain on each other's syllable rates. By contrast, [2] claims that the basis for the inter-speaker coordination are sequences of beats.

While these models have been proposed to account primarily for cases of speaker changes without overlap, they should remain valid for changes accompanied by overlapped speech. In an earlier study we have demonstrated that overlap initiations are more frequent around the boundaries of the first overlapped syllable than at other locations within the syllable [3]. In this paper we investigate whether similar evidence of coordination can be found for the inter-stress interval (ISI).

2. METHOD

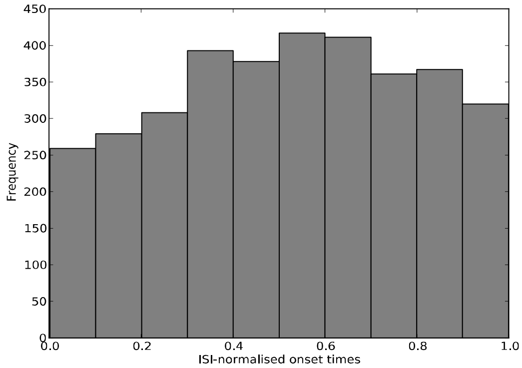
Stretches of overlapping speech in the Switchboard corpus were derived from inter-pausal units (IPU) bounded by at least 0.1 s of silence. For each overlap, the ISI during which the overlap was initiated was identified. The overlap onset was then normalised relative to the duration of this first overlapped ISI: *ISI-normalized onset time* was calculated by dividing the duration of the interval from the onset of the overlapped ISI to the onset of the overlapping utterance by the duration of the overlapped ISI. Overlaps coinciding with the first or last ISI in overlappee's IPUs were excluded from the analysis. Overall 3439 overlaps were analysed.

ISI boundaries were calculated from automatic phone segmentation, lexical stress labels, POS tags and orthographic transcriptions. Due to the inaccuracies involved in this method and the resulting problems with finding exact vowel onsets, ISI boundaries were set at half of the duration of the stressed vowels.

3. RESULTS

Figure 1 shows the distribution of ISI-normalised onset times for all overlaps in the corpus. One sample Kolmogorov-Smirnov test revealed that the distribution is significantly different from the uniform (flat) distribution ($p < 0.001$). The distribution reaches a maximum around the value of 0.5, i.e., in the middle of the ISI.

Figure 1: ISI normalized onset time distribution



4. DISCUSSION

The fact that the distribution in Figure 1 is significantly different from uniform indicates that the timing of overlap initiations is non-random. Specifically, the peak around 0.5 suggests that dialogue participants are most likely to start speaking in overlap in the middle of the interval between successive stressed vowels of their dialogue partner.

Along with the results reported in [3], the non-randomness of overlap initiations provides evidence that dialogue participants have access to the rhythmic structure of their partner's speech both at the syllabic and the ISI levels.

5. REFERENCES

[1] E. Couper-Kuhlen. 1993. *English speech rhythm: form and function in everyday verbal interactions.* Amsterdam: John Benjamins.

[2] M. Wilson and T. P. Wilson. 2005. An oscillator model of the timing in turn taking. *Psychonomic Bulletin and Review*, vol. 12, n. 6, pp. 957-968.

[3] M. Włodarczak, J. Šimko and P. Wagner. Syllable boundary effect: temporal entrainment in overlapped speech. *Proceedings of Speech Prosody 2012*, pp. 161-164.

Phrase-final lengthening, focus and Australian-English vowel length contrast

Ivan Yuen, Felicity Cox, Katherine Demuth

Department of Linguistics, Macquarie University, Australia
ivan.yuen@mq.edu.au

Keywords: phrase-final lengthening, focus, vowel length contrast

1. INTRODUCTION

Speech is organised into phrasal units to facilitate communication. Duration can be used as a cue to signal the boundary of a phrase, as in phrase-final lengthening, or to highlight a constituent as in focus. However, duration is also used to signal phonemic contrasts such as phonemic vowel length. Boundary-related lengthening and focus could therefore interact with vowel length distinctions at the segmental level. The goal of the current study is therefore to examine how phrase-final lengthening is distributed over a monosyllabic word, and the extent to which it interacts with phonemic vowel length contrasts, as found in Australian English [2].

Recent studies of phrase-final lengthening in American English show progressive final lengthening: Phrase-final codas were found to be lengthened the most, followed by vowel nuclei [1,4]. On the basis of these studies, we predicted that the various units of the phrase-final monosyllable (onset, nucleus, coda) would be longer than their phrase-medial counterparts, and that the magnitude of lengthening would diminish from the coda to the onset. We also predicted lengthening as a function of focus.

2. METHOD

The current study employed an imitation task to elicit data from ten (3M,7F) speakers of Australian English with a mean age of 25. Sixteen 3 word utterance prompts (e.g. *Now X goes*) were pre-recorded by a female speaker of Australian English from Sydney. The target words were nonce words contrasting in vowel length (/kɐs/-/kɐːs/, /gɐs/-/gɐːs/). These appeared in three experimental conditions: i) short versus long vowels. ii) phrase-final versus phrase-medial positions, and iii) focus versus no focus. The stimuli were played through two loudspeakers. Each participant was instructed to repeat the 16 auditory prompts (each varying by onset voicing /k g/, vowel length, phrasal position, focus). The responses were recorded onto a computer at 22KHz through ProTools LE for acoustic analysis.

3. RESULTS & DISCUSSION

Repeated measures ANOVA were conducted on the stop burst duration, vowel duration, fricative coda duration, rhyme duration, and nucleus/rhyme ratio. As predicted, the results confirmed progressive phrase-final lengthening. Overall, phrase-final position, long vowels and focus all contributed to an increase in rhyme duration. Specifically, burst durations in the onsets were longer in the focus condition than no focus condition. For the nucleus, the three factors all contributed to increased duration. Additionally, focus interacted with vowel length by lengthening long vowels more than short vowels. For the coda, duration was longer in the phrase-final focus condition than the phrase-medial focus condition. Consistent with [3], coda duration was also found to be longer after short vowels than long vowels.

Interestingly, both phrasal position and vowel length affected the nucleus/rhyme ratio, but not focus, suggesting the *rhyme* to be the domain for accentual lengthening. The two factors also interacted so that the vowel length contrast was greater in medial position. The results are discussed in terms of multiple processing units/domains in speech planning.

4. REFERENCES

[1] Berkovits, R. 1994. Durational effects in final lengthening, gapping and contrastive stress. *Language and Speech* 37, 237-250.

[2] Cox, F. 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. Australian Journal of Linguistics, 26, 147-179

[3] Cox, F., Palethorpe, S. 2011. Timing differences in the VC rhyme of Standard Australian English and Lebanese Australian English. Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong.

[4] Turk, A., Shattuck-Hufnagel, S. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, 397-440.

Acknowledgements: This research was funded by Macquarie University, and the grant: NICHDR01HD057606 (Demuth & Shattuck-Hufnagel).

Timing of tongue movement in preadolescents

Natalia Zharkova

Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, UK
nzharkova@qmu.ac.uk

Keywords: timing, coarticulation, tongue, ultrasound, dynamics

1. INTRODUCTION

This study compared patterns of tongue movement during the consonant /s/ in preadolescents and adults. Our earlier study of anticipatory lingual coarticulation in /sa/ and /si/ syllables showed that the vowel coarticulatory effect was present from the consonant onset for adults, while for preadolescents it was only apparent later in the consonant ([3]). The coarticulatory effect was judged to have occurred when the consonant tongue shape in the context of /a/ was significantly different from that in the context of /i/. The tongue travelled in the midsagittal plane significantly more during /s/ before /i/ than before /a/, with the blade moving forwards and the root moving upwards towards /i/ throughout /s/, producing the same "pivot pattern" ([1]) in both groups of speakers. The present study used data from a larger number of speakers, comparing tongue displacement over the consonant within vowel context across age groups, in order to better understand the dynamics of coarticulation.

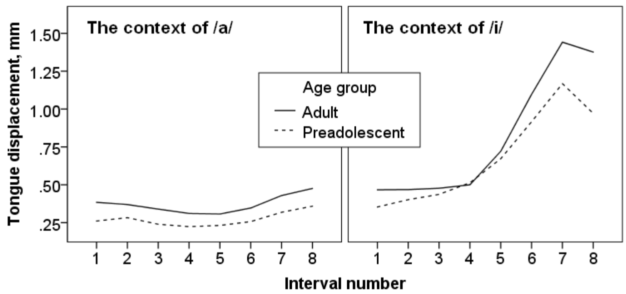
2. METHOD

The same dataset was used as in [3]. The speakers were fifteen preadolescents (aged between 10 and 12 years old) and fifteen adults, native speakers of Standard Scottish English. The stimuli were the syllables /si/ and /sa/, in a carrier phrase, with six repetitions of every target. The ultrasound signal, at 100 Hz, was synchronised with the acoustic recording. In each token, tongue curves were traced for every ultrasound frame during the consonant. Distances between each consecutive pair of nine equally spaced curves were computed using the nearest neighbour method ([2]). Repeated measures ANOVAs were performed for each vowel context, comparing distances across Interval and Age Group.

3. RESULTS

The results on tongue displacement are presented in Fig. 1. In the context of /i/ the preadolescents had relatively more movement in the first half of the consonant than the adults. The ANOVA for /s/ from /si/ showed a significant interaction of Interval and Age Group: $F(2.31, 64.62) = 4.803$, $p = 0.008$. The interaction was not significant for /s/ from /sa/. The age-related difference in tongue movement over /s/ in the context of /i/ can be represented by ratios of the sum of displacements in the first half of the consonant to the sum of displacements in the second half of the consonant. For /s/ from /si/, this ratio was 0.46 in the preadolescents, and 0.41 in the adults.

Figure 1: Tongue displacement in consecutive intervals over /s/.



4. DISCUSSION

The observed age-related difference in tongue dynamics could go some way to explain the later coarticulation onset in preadolescents than in adults, reported in [3]. Relatively more tongue movement during the first half of /s/ from /si/ may be a strategy used by preadolescents to compensate for the late onset of the vowel coarticulatory effect. This interpretation would suggest that preadolescents may require more time than adults for planning and articulating a combination of segments that involves a substantial amount of tongue movement.

5. REFERENCES

[1] Iskarous, K. 2005. Patterns of tongue movement. *Journal of Phonetics* 33, 363–381.

[2] Zharkova, N. & Hewlett, N. 2009. Measuring lingual coarticulation from midsagittal tongue contours: description and example calculations using English /τ/ and /A/. *Journal of Phonetics* 37, 248–256.

[3] Zharkova, N., Hewlett, N., Lickley, R., Hardcastle, W. 2011. Lingual coarticulation dynamics in preadolescents: an ultrasound study. Poster at the International Child Phonology Conference, York, UK, 16–18 June 2011.