WAVES, BEATS AND EXPECTANCY

Eric Keller

IMM, Lettres, University of Lausanne
eric.keller@unil.ch

ABSTRACT

In speech, “beats” are found at rapid onsets or at strong increases of low-frequency components of voicing (typically 200-800 Hz). These sites correspond to well-documented foci of perceptual attention (e.g., P-centres). Given the precise coordinated action they require, they probably necessitate exceptional articulatory coordination. The on- and offsets of the repeated short sentences with characteristic durations in studies by Port and colleagues were marked by strong beats, and in studies on continuous French and English sentences, strong beat positions were places where timing agreement between speakers was greater than in weak beat positions. Beats thus enjoy privileged status in the phonetic chain.

In this study, a mechanism for the emergence of beats is proposed. Examples from physics (e.g., pendulum, sine tones, or water waves) indicate that beats are produced by interferential patterns between similar events, creating mutually reinforced wave forms (“beats”) alternating with weakened wave forms (“off-beats” or “anti-beats”). In biological “coordinative structures”, similar task requirements can also create interferential events that translate into beat, e.g. the need to coordinate the horse’s anterior and posterior body portions for a canter. In speech, it can be argued that strong voice onsets coincide with a neurological anticipation of such onsets, which is likely to lead to the creation of beat reinforcement. Weakening would be predicted for rapid subsequent events. The notion of beat patterns operating within coordinate structures promises a number of useful hypotheses for the temporal structuring of gestures in speech.

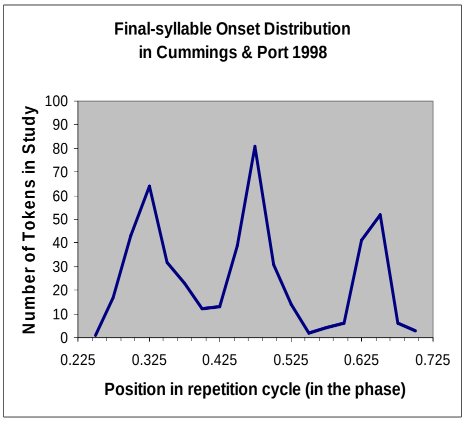
Keywords: beat, rhythm, timing, coordinative structures.

1. INTRODUCTION

Like many others, I originally expected that perceived beat in speech could easily be related to durational measures. When hearing the nursery rhyme “**Pet**er, **Pet**er **pump**kin-**eat**er, had a **wif**e and **couldn't** keep her”, I expected to find systematic relationships between the durations connecting the five stressed syllables. But Lehiste’s detailed refutation of isochrony [10] suggested otherwise. It summarized research showing that listeners judged inter-stress intervals or syllable durations to be isochronous, even if they were far from equal in duration. This led to the conclusion that such timing events probably lay below the threshold of durational perception, and that for some cognitive reason, the “ear jumped to synchronic interpretation”.

This argument proved difficult to dislodge, but in 1998, Fred Cummins and Robert Port published their paper showing the famous triple peaks for durations of “dig for a duck” type of sentences [2]. In this repetition experiment, English-speaking subjects were induced to produce durations that should theoretically have ranged anywhere between very fast and very slow.

Figure 1. Data recreated from Figure 1 in Port (2003) [12]. Final-syllable onset distribution as a phase angle in the repetition cycle created by repeating a 4-syllable phrase like “Dig for a duck” in time with two tones. The first tone marked the onset of the phrase and the second tone varied randomly from 0.20 to 0.80 of the phrase’s duration.



The subjects showed however a distinct preference for durations approximately one half, one third or two thirds of the duration of the overall repetition cycle. These experiments suggested that in some speech conditions, speech events organize themselves into measurable durational regularity, although the regularity had to be expressed as a percentage of an overall duration. The harmonic durations (33%, 50% and 66% of the repetition cycle) suggested the effects of “coordinative structuring” of the speech event [11, 12] (for “coordinative structures”, see [5, 16]). These experiments raised a number of questions:

(1) Were these results limited to the experiment’s artificial repetition design, or would the harmonic concept generalize to continuous speech?

(2) The on- and offsets of these sentences were marked by strong beats, and sentence durations were measured by the distance between these beats[1]. Did the beats have a temporal organizing effect?

Two experiments were run to examine these questions. They were reported in [7, 8] and are summarized here.

2. GREATER PRECISION AT BEATS

In experiment 1, nine French speakers and in experiment 2, two English speakers read the same set of prepared sentences for each language. In this study, “phrases” were considered to be uninterrupted and well-articulated stretches of speech. Of initial interest were the positions of beats within the phrase. According to the harmonicity hypothesis, it was expected that beats would cluster around harmonic positions within the phrase, in particular strong beats (those with strong onset slopes). The results were weak; only at around 50% of the phrase, a certain clustering appeared (Figure 2). Results for the English speakers were similar.

It was then examined if strong beats induced more temporal stability than weak beats. This was measured by the average time difference that the various speakers showed for the same beat positions within the phrase (expressed as a percent of the total phrase). The results were significant when the effects of very strong beats were compared with very weak beats (right edge of Figure 3). In those comparisons, speakers showed about 0.5% more temporal agreement on strong beat positions than on weak beat positions. Average agreement for beat positions was 2.27%. Differences reached the statistical significance level of $p < .05$ at condition 7, and the greatest difference (condition 8) was different at $p = 0.0175$ (Paired t-Test, Mean of Paired Differences = 0.462, t = 2.49 w/36 df).

Beats thus have a temporal structuring effect in both artificial repetition experiments (like those of Port and colleagues) and in continuous speech (like mine). We thus need to define the theoretical basis of “beats”. Where do beats come from?

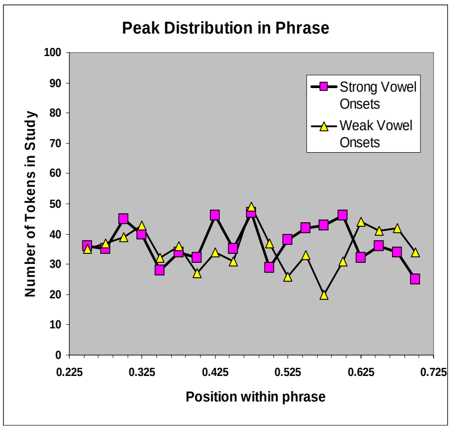


Figure 2. Data from the experiment with French subjects (*N* strong beats = 703, *N* weak beats = 671). There is minor clustering of beats at about 50% of phrase duration. The English speakers showed similar results.

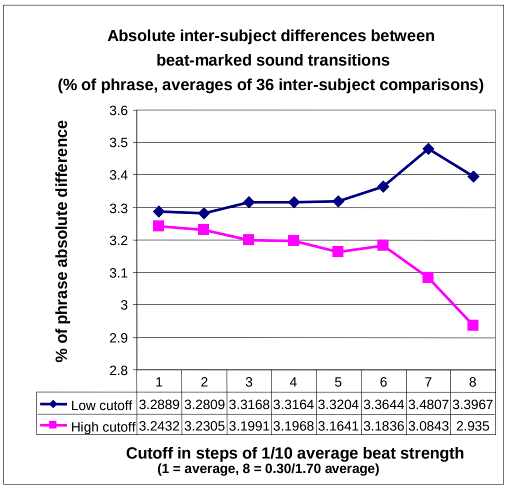


Figure 3. Comparisons between French speakers. The data cut-offs move from average cut-off (left) to high/low cut-offs (right). At the *left extreme*, inter-subject variation is calculated for all parts of data above and below the beat average, and at the *right extreme*, variation is calculated only for very strong and very weak beats; intermediate data is eliminated. Similar results were found for the English speakers.

---

[1] The exact definition of a beat varies. In [7, 8], the definition used in Port (2003) [12], derived from [3] was used: “the beat location can be approximated ... by measuring the amount of energy in lower frequencies (between 200 and 800 Hz), smoothing sufficiently and then looking for large energy onsets”.

3. THE BEATS HYPOTHESIS

Table 1 lists a number of “beats” in physical events, in objects and in animals. In every case it can be seen that *beat results from interference between two similar, but slightly different events.*

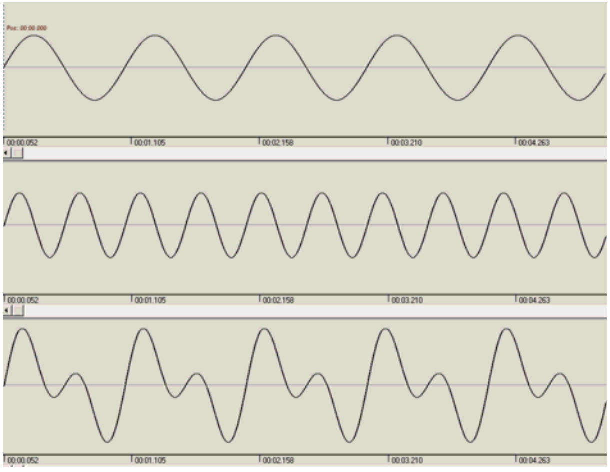
The classic example is the linear addition of two slightly different sine waves. The summing of two sine waves with different frequencies (F1 and F2, where F1<F2) produces a larger complex wave that fluctuates at rate F2-F1 (Figure 4). Two simple steady wave forms thus combine to form an *interferential pattern* of stronger and weaker wave forms. The stronger wave form is ordinarily called a “beat”, and the terms “off-beat” or “anti-beat” are sometimes used to describe the weaker wave form. Strong and weak beats thus produce a combined larger cyclical temporal entity[1].

Table 1: Some examples of beat

	Waves	Beats
Sound waves	Simple sine wave: heard as a single tone consisting of the repetition of the same sine waveform	The addition of two sine tones with similar frequencies produces a sequence of alternating strong and weak wave forms
Pendulum	Balanced action: impetus applying approximately equal force in both directions	Unbalanced “limping” action: impetus applies force unequally
Animal and human gait	Normal biped walk (e.g., ostriches, spontaneous human walk) produces an equilibrated gait	Canter in a horse: an unequilibrated gait consisting of three minor and one major hoof beats

The comparison of a balanced and a “limping” pendulum illustrates another interferential pattern, opposing in this case pendulum impetus and gravity. In an equilibrated clock, impetus counteracts gravity more or less equally in both directions, while in a sideways hanging clock, impetus encounters stronger gravity effects on one than on the other side of the swing. The net effect is a temporal deregulation between the two directions of the swing and the creation of a distinct temporal entity (a “ripple”) consisting of a beat and an off-beat movement.

Figure 4. Beats obtained by adding two waves of different frequencies



The principle of beats emerging from interferential wave patterns probably extends to movements observable in more complex articulated bodies. An interesting example is canter in a horse. Canter is situated between walking and galloping, and usually consists of the right front foot leading the set of four hooves hitting the ground in sequence. This pattern probably emerges because the front and the back portions of the horse coordinate a rapidly occurring action in an articulated structure of bone, muscle and fatty tissue[2]. However when the movement is not as rapid as in a walk or in a trot, a balanced step usually emerges between the four legs, unless a limp is provoked by injury or an unevenly placed horseshoe.

In all these cases of physical phenomena, articulating objects or biological coordinate structures, a “beat” emerges from interference between two sources that are in sufficient temporal and physical proximity. This suggests that the origin of beats in speech might be found in inferential patterns involving linguistic material

---

[1] It is tempting to call this entity a “foot” (as in “trochaic foot”), but that would jump the issue of the relationship between “beat” and “accent” which is unlikely to be straightforward (see below). Instead, I would like to propose the term “ripple”: a “ripple” consists of a beat and an off-beat.

[2] An illustrative moving-gif animation of a horse’s canter can be seen in Wikipedia at page http://en.wikipedia.org/wiki/Canter.

that is sufficiently similar and occurs in sufficient temporal proximity.

One might hastily propose at this point that phonetic beat result from interference between two acoustic wave forms. However, simple acoustic wave forms are not good candidates for such interference, since they would have to be excessively similar, which would exclude beat patterns establishing themselves between onsets to different vowels.

However a more complex form of interference does appear compatible with a number of well-known results in phonetics, psycholinguistics and neurophysiology. In short, it will be proposed that *beat-marked voice onset positions emerge from the coincidence between a rapidly decaying neurological anticipation of a critical voice onset event and the actually occurring event.* It is argued that if a critical timing event occurs at the expected time, the expected event is neurologically added to the actually occurring event and manifests itself externally as a stronger phonetic event. Immediately following events would be depressed as part of the same neurological expectancy pattern, and events that occur at unexpected places in the phonetic chain would be less exposed to beat formation.

4. NEURAL EXPECTANCY OF VOICE ONSETS

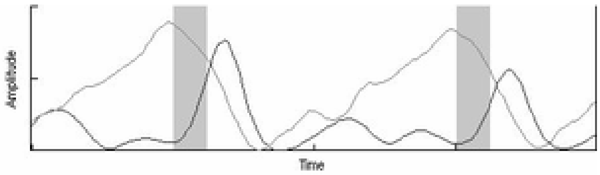
This hypothesis exploits well-established notions of expectancy of time-critical events, such as notes in music or voice-onset times. The argument is not new. In 1992, Desain proposed a method for constructing mathematical functions for the various elements contributing to the expectancy of event onsets in music [4]. This model accounted for a great number of experimental results in rhythm detection and in the memory of auditory events that were then known. In this model, the complex temporal patterns that form a listener's expectation of the same event in the future were described as a set of simple anticipation curves based on previous experiences of musical events (previous musical notes occurring at specific delays). When added together, these curves form the complete and complex expectancy of coming musical events.

The model predicts that very short and very long durations would be difficult to judge since on the one hand, there is a rapid fall-off of expectancy, with little expectancy for events greater than about 1 second, and on the other hand, there are no predictive instances for very short durations. On the other hand, perceptual judgement should be best and expectancy should be the greatest for about 600 ms.

Largely in line with this basic model, we hypothesize that speaker-listeners maintain a neurological planning stage of less than one second during which time the anticipation of regularly occurring punctual speech events (such as strong voice onsets) is combined with actually occurring articulatory and/or acoustic events of the same type, in order to form the basis of beat. Also in line with Desain's model, we shall initially assume that the model is essentially additive in nature. That is, actual events coinciding with predicted events will essentially sum in linear fashion, and events that do not coincide will be submitted to essentially linear reduction due to the absence of a corresponding predicted event. The degree of linearity of the prediction – just as the entire model – will have to be experimentally verified.

Recent neurophysiological experiments have contributed support to this model. For example, Snyder and Large [14, 15] examined auditory rhythmic anticipation with two types of gamma band (20-60 Hz) analysis of EEG recordings taken when metronome beats were presented. In one analysis called "induced" (grey line in fig. 5), they measured the phase-unrelated amplitude of the gamma band activity before performing averages over trials; in the other analysis ("evoked", black line in fig. 5), they time-locked the trials to the stimulus onset, averaged the trials by taking into account stimulus-related phase, and then took the amplitude. Of interest here is the induced analysis which shows an anticipatory time course that is quite similar to the well-known "readiness potential" measured contra-laterally prior to movement onset in distal limbs (see [9] for initial publication).

Figure 5. Gamma-band activity prior and subsequent to the presentation of a metronome beat. Adapted from an illustration appearing on E.W. Large's web page www.ccs.fau.edu/~large/Publications/SnyderLarge2005.html.



Another paradigm that permits us to assess neural expectancy is the omission paradigm giving rise to “emitted potentials”. These potentials (“emitted P300 potentials”) are positive peaks occurring about 300 ms after the point of where absence is noted in the stimulus [13]. Potentials of this type are well-described and respond to similar environmental modification as do potentials evoked by actual events [15]; furthermore, they do not depend on automatic evocation, since they can also be generated as part of an imagined tone series [6].

In brief, the expectancy hypothesis seems plausible, and can readily be tested in perception[1]: it is possible that expected plosions created by previous strong voice onset/increases in similar linguistic contexts will augment the P300 generated by the real event, while a similar P300 in an immediately following or unexpected site should be depressed, because they are situated in the off-beat or unexpected position of the expectancy curve. A similar prediction can be made for articulation: it is possible that the expectation of a strong voice onset/increase combine with the augmented potentials needed for the production of a precise motor event. In such a combination, the expected and the realized events would produce an externally augmented beat, and would contribute to the diminution of immediately subsequent beats within a short time period (of less than one second). Unexpected beats outside of this time period would receive no support from expectancy.

5. CONCLUSION

We have described beats as objectively measurable aspects of the speech chain. Strong beats typically mark voice onset times that are known to have great perceptual and articulatory salience. Speakers of French and English have been found to show more temporal agreement for strong than for weak beats.

A mechanism for the emergence of beats was proposed here. Examples from physics (e.g., sine tones, water waves, pendulum) indicate that beats represent interferential patterns between similar events that cause mutually reinforced wave forms (“beats”) as well as weakened wave forms (“off-beats” or “anti-beats”) within a certain time period. It was argued that in rapidly moving articulated biological structures (so-called “coordinative structures”), similar task requirements also create interferential events, as can be seen in the need to coordinate the horse’s anterior and posterior portions in a canter.

In speech, it was proposed that strongly beat-marked voice onset positions emerge from the coincidence between a rapidly decaying neurological *anticipation* of a voice onset event and the *actually occurring event.* For rapidly occurring events occurring inside of the expectancy window but outside of expected beats, weakening is predicted.

It is argued that the sharp temporal precision required for the production of a clear voice onset event requires increased attention in the movement preparation mechanism. Also, a similar increased expectancy is created for perceiving a strong voice onset event in these locations. These patterns thus serve to provide temporal “anchor points” within the speech chain and they would explain the greater temporal agreement between speakers for these locations in texts read aloud. The temporal abruptness of beats set them apart from other speech features: few features of the continuous speech chain (if any) are as perceptually salient as clear voice onsets. This makes clearly marked voice onsets ideal candidates for temporal structuring, in any language.

A number of predictions are made by this hypothesis. For example, such interferential patterns in speech can be seen as stimulating the emergence of alternating and syncopating accentual (stress) patterns in rapidly produced speech, as would be illustrated by the large distribution of such patterns in the world’s languages (for English, see [1]).

A question that needs rapid clarification concerns the relationship between beats as defined as strong initiations of voiced portions of speech, and historically formed accentual patterns in a given language. The relationship between the two types of prominence is unlikely to be straightforward because of historical change that shaped the precise form of contemporary accentual patterns. Even if beats played a role in the emergence of accents, language change is likely to have modified accentuation over time, and this may set it apart from beat-induced sequences. Beats are best seen as real-time phonetic events

---

[1] Contamination from strong myographic potentials from the oral region renders a neurophysiological verification considerably more difficult for articulation.

emerging from a long term memory trace of phonetically similar events, while accent is a stabilized shared phonetic feature of the community of speakers of a given language. The two may well show only partial overlap, particularly in some languages.

At the same time, certain effects are predicted for beats as well as for the historically shaped accentual patterns. For example, we expect increased differentiation between beats (or accents) situated within the expectancy window, while the differentiation should decrease when beats/accents are outside the expectancy window.

Another question concerns long-term memory for expectancy. Hypothesized and demonstrated expectancy curves decay over the short term (within one second), and yet the French and English speakers showed extensive agreement on which syllables they chose for beat position[1], and as was said, they showed greater temporal agreement for beat position. The exploitation of long-term memory was not directly incorporated into Desain's model of the anticipation of musical events, yet seems to suggest itself very powerfully for the linguistic context. If beats are indeed produced as presented here, the results for phonetic beats can only be explained if one supposes long-term retention and the sharing of typical temporal positions for beats within a linguistic community.

Finally, it is possible that the repeated and temporally constrained neuronal expectancy of beats may contribute to encouraging pulsing patterns in speech within internal coordinative structures. If so, it may open another perspective on the possible neural bases for the proposed oscillatory behaviours of speech timing.

6. ACKNOWLEDGEMENTS

Part of this research was supported by work conducted under the European COST 258, 277 and COST 2102 projects, in part co-financed by the Staatssekretariat für Bildung und Forschung (SBF), Berne, Switzerland.

7. REFERENCES

Many publications and previous talks by the author can be found via http://www.unil.ch/imm/page22063.html.

[1] Clopper, C.G. (2002). Frequency of Stress Patterns in English: A Computational Analysis. IULC Working Papers Online, 2:2. https://www.indiana.edu/~iulcwp/abstracts.cgi?which=2.

[2] Cummins, F., & Port, R. (1998). Rhythmic constraints on speech timing. *Journal of Phonetics, 26,* 145-171. http://www.asel.udel.edu/icslp/cdrom/vol4/437/a437.pdf

[3] Delgutte, B., & Kiang, N. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *Journal of Acoustical Society of America, 75,* 866–878.

[4] Desain, P (1992). A (De)Composable Theory of Rhythm Perception. *Music Perception. 9(4).* 439-454.

[5] Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51,* 347–356.

[6] Janata, P. (2001). Brain Electrical Activity Evoked by Mental Formation of Auditory Expectations and Images. *Brain Topography, 13:3,* 169-193.

[7] Keller, E. (in press). Beats for individual timing variation. In A. Esposito, E. Keller, M. Marinaro, M. Bratanic (eds.), *The Fundamentals of Verbal and Non-Verbal Communication and the Biometrical Issue.* IOS Press.

[8] Keller, E. (2006). *Individual speech rhythm variation within the plosive structure of speech,* Part 1: The beat structure for speech timing, Part 2: Verification of Hypotheses. Nato Advanced Study Institute Advanced Study Institute (Asi) - Summer School "E. R. Caianiello" on The Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue, September 2 - 12, 2006 - Vietri sul Mare, Italy.

[9] Kornhuber, H.H., & Deecke, L. (1964). Hirnpotentialänderungen beim Menschen vor und nach Willkürbewegungen, dargestellt mit Magnetbandspeicherung und Rückwärtsanalyse. Pflügers Arch Eur J Physiologie 281: 52.

[10] Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5,* 253-263.

[11] Port, R., Tajima, K., and Cummins, F. (1999). Speech and rhythmic behavior. In Savelsburgh, G. J. P., van der Maas, H., and van Geert, P. C. L., eds., *The Non-linear Analysis of Developmental Processes.* Elsevier, Amsterdam.

[12] Port, R.F. (2003). Meter and speech. *Journal of Phonetics, 31,* 599-611.

[13] Ruchkin, D.S., Sutton, S. & Tueting, P. (1975). Emitted and Evoked P300 Potentials and Variation in Stimulus Probability. *Psychophysiology 12 (5),* 591–595.

[14] Snyder, J. & Large, E. W. (2002). *Neurophysiological correlates of meter perception.* In C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition.* Adelaide: Causal Productions. www.ccs.fau.edu/~large/Publications/publications.html

[15] Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24,* 117-126. www.ccs.fau.edu/~large/Publications/publications.html

[16] Tuller, B., & Kelso, S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance VIII* (pp.429–451). London: Academic Press.

---

[1] One male and one female speaker chosen randomly from the nine French speakers used the same 36 strong beat syllables out of a total of 38 syllables that showed strong beats for either speaker in sentence 1 of the reading task. Here: strong beats are those > median [8].