

The social meaning of tempo

Tyler Schnoebelen

Contents

Introduction	2
What is tempo?	2
Some meanings of tempo	3
Emotions and arousal.....	5
Time ideology.....	9
Context, interaction, performance, Captain Kirk.....	10
On acting	11
Starburst: The William Shatner story	12
Case 1: Variation in a monologue	14
Case 2: Superburst.....	17
Case 3: A woman inside Captain Kirk	22
Conclusion and next steps	30
Works cited.....	32
Appendix A: Slow talk indexical field	35
Appendix B: Indexical fields as lists (with references and some notes)	36
Fast speech.....	36
Slow speech.....	37
Appendix C: Survey results	38
Demographics.....	41
Appendix D: Music and emotion.....	43

Introduction

This is an exploration of the social meaning of tempo. Speech tempo can be deployed consciously to achieve particular effects, but it can also send cues to listeners that the speaker didn't intend to convey. What this means for us is that tempo is a stylistic resource, expressive for both speakers and listeners—creating expectations about the speaker and their attitude toward a situation and audience. These expressions, however, don't happen in a vacuum. The use of tempo reflects and constructs various ideologies, allowing tempo to signal emotional state, occupation, geographical origin, ethnic identification, and more.

Because tempo is relatively easy to measure, it has been evaluated by researchers looking for acoustic reflexes for emotions, competence, extroversion, and a whole host of other characteristics. Much of this work seems stuck, probably because it lacks an adequate theoretical foundation. Often the studies are merely about detection and assume a high degree of correspondence between what's going on in the speaker's head and what's happening in the acoustic signal. So to begin, it will be necessary to look skeptically at research goals. Scherer is quite articulate about the paradox that researchers face in looking at emotion in speech.

Whereas judges seem to be rather accurate in decoding emotional meaning from vocal cues, researchers in psychoacoustics and psychophonetics have so far been unable to identify a set of vocal indicators that reliably differentiate a number of discrete emotions. (Scherer 1986: 143-144)

Of course it's true that some emotions and social categories may be better conveyed by non-speech signals, but even for those that are detectable, researchers neglect the fact that speech is not merely reflective of internal states. Rather, speech is situated in particular contexts, between a speaker and an audience, having particular effects on the audience, both unintended and intended.

In this paper, I attempt to build out indexical fields for tempo. I sketch how it is that different aspects of the fields adhere, positing that the two major themes have to do with (i) arousal, and (ii) perspectives on time. Those discussions tend to be about fast and slow speech, but to demonstrate some of the variations that are possible beyond fast/slow, I offer a close analysis of several scenes in *Star Trek*. These scenes also allow me to show the contextual/interactional nature of style—and tempo in particular.

What is tempo?

The easiest way to operationalize speech tempo is to count the number of syllables in a given time period and report syllables per second. This is known as the **speech rate** and in English it usually runs between 2.3 and 4.3 syll/sec (Goldman-Eisler 1968).¹ It is possible to measure an overall speech rate for a person, and if we control for the types of situations we have recordings of, then we could report that Robin speaks faster than Bernice, who speaks faster than Harry; they all speak faster than the average person from Orlando, but slower than the average inhabitant of Miami.

Such a report may be fascinating to Bernice and her pals and the geographical facts may make a nice blurb in the local Floridian papers. But to say something that's actually interesting, we need more than a blanched average. We need to examine how speech rates vary. When does Robin speed up? Is it similar

¹ It can reach into the realm of 6 or 7 syll/sec. Comedians whose shtick is fast talking can sustain 6.7 syll/sec, but that sort of rate only sounds natural when it happens for brief moments.

to when Bernice speeds up? It is by identifying patterns of variation that we start to figure out the social meaning of tempo.

I fall back to talking about “tempo” because I think “speech rate” focuses too much on an average. In this paper, I will report speech rates for small segments of speech so that I can trace their acceleration and deceleration across intonational phrases and speech turns. But even within a given intonational phrase, I will point out how the syllables fall relative to one another—where there are pauses, elongations, staccato bursts and so forth. I use the word “tempo”, in part, because I will not just be reporting speech rates, but also looking for evenness and irregularity. This is relatively easy to visualize and I have been experimenting with ways to measure the degree of “burstiness”, too. In this expansion of tempo beyond speech rate, I begin to approach rhythm. Tempo is certainly a component of **rhythm**, but rhythm also requires us to tackle issues of stress/accent—something outside the scope of this paper.

The nice thing about an exploratory paper is that it’s permissible to be provisional in some areas. And I feel rather provisional in my definition of tempo. One could read that and think that I am building a house with a foundation of Jell-O, but that isn’t the case. A gelatin foundation implies a lack of precision, but my working definition of tempo is quite specific and measurable. The problem is that there seems to be something *extra* that is missing. In music, a composer adds a tempo marking not just to set the rate, but to set the intensity and the feeling. “Intensity” or “feeling” may be the specters that haunt my halls, but I do not attempt to add them here.²

Some meanings of tempo

Tempo doesn’t mean one thing; nor can we say that “fast tempo” means something particular. A fast tempo has multiple meanings. Depending upon what other variables it combines with, it can mean happy, New Yorker, angry, con artist, and a number of other things. Its meaning is indeterminate and requires other cues—some of which are linguistic, others which are not. This doesn’t stop people from explicitly commenting on what “fast talkers” are like or what kinds of people talk fast.

Stylistic variables are often explicitly commented upon, so scanning corpora and surveying people are good ways to start charting meanings. In the case of tempo, we can also take advantage of research that has been published in psychology, phonetics, and musicology.³ While it’s only become possible to measure some acoustic features rather recently, tempo is something that researchers have been able to measure for quite some time. In a survey of about 40 papers, I have found that most researchers on tempo have been primarily interested in detecting a correlation between speech rates and perceptions of particular qualities, but they don’t spend much time imagining how the link comes about or is maintained. Occasionally “sociocultural factors” will be mentioned, but only in a single sentence that says, “By the way, sociocultural factors may be part of the story.” My work here is to unpack how meanings get assigned to tempo and how those meanings are connected. These connections are ideological in nature and have social consequences: how you time your speech tells people something about the kind of person you are and how you regard the situation. As an important first step, I have tried

² As an aside, it is not obvious how much we want to pull from the musical definition of tempo, though it should clearly serve as inspiration and counterpoint. For what it’s worth, even in research about emotion in music, people generally define tempo only as beats-per-minute.

³ There doesn’t really seem to be much in the sociolinguistic literature. Ray and Zahn (1999) is a good resource, though.

to build an indexical field—a “constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable. The field is fluid, and each new activation has the potential to change the field by building on ideological connections” (Eckert 2008: 453).

The 40-odd items that you see in Figure 1, demonstrate the indexical field for fast talking. These attributes come from how people in several corpora talk about fast speech⁴ and from research by psychologists and others who have found statistically significant relationships between, say, fast speech and listeners’ perceptions of intelligence. It is augmented by a pilot survey I ran on 50 people, asking what sorts of people talk fast.⁵

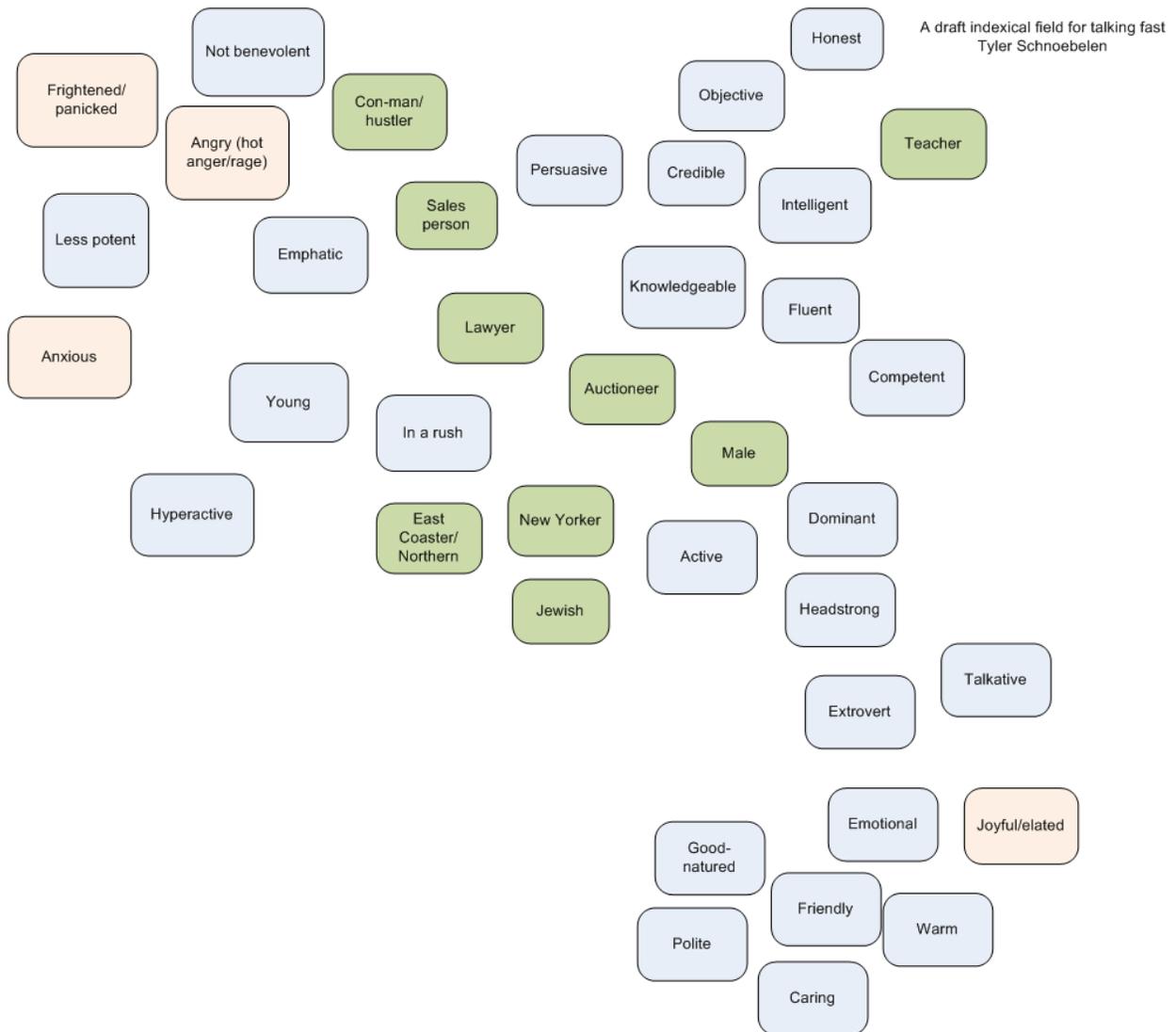


Figure 1: Indexical field of fast speech tempo, see also “Appendix A: Slow talk indexical field”.

⁴ For example: “And I think people have a certain perception about Jews, you know. They’re fast-talking. They’re -- you know, they’re, you know, fraught with neuroses and all this kind of thing.” Adam Goldberg, Fresh Air interview, NPR, 10/27/2005.

⁵ See “Appendix C: Survey results” for more information on the survey; see also “Appendix A: Slow talk indexical field” and “Appendix B: Indexical fields as lists (with references and some notes)”.

If these characteristics were presented as a list, we might be distracted by all the contradictions—how can fast speech be about both rage and joy? How does it signal an honest person and a con-man? Laying out the characteristics in a field, however, allows clusters to form. Some of these we sense intuitively—in America, there is a cultural association between “New Yorker” and “Jewish”, for example, which can be traced both to historical settlement patterns and current demographics, but which also work out to be ideological judgments about shared outlook and behavioral patterns. The connection between New Yorkers and other elements are also relatively straight-forward. As the stereotype goes, New Yorkers are in a hurry. Portrayals of New Yorkers often show them to be aggressive—an attribute that connects to “active” and “emphatic”, and which can be read as a version of “angry”, too.

The cluster in the upper right has to do with credibility and honesty. Clearly, honesty is connected to objectivity and credibility. These characteristics also make someone persuasive—which is enhanced by a perception of knowledgeable and intelligence, too. Certain occupations require persuasion—lawyers and salespeople, for example. But the very fact that talking fast is persuasive can also cause people to be circumspect about it. We know that con-men want to persuade us, too.⁶

Even when speech rates are held constant, men are judged as speaking faster. And, indeed, women have been found to slow down their speech as mixed-gender interactions progress (Kimble et al 1981: 1051). But associating men with fast talking is fundamentally ideological as it has so much to do with other attributes men are associated with: dominance, for example, and other characteristics in the “dynamic” cluster (active, extrovert, talkative, etc).

The indexical field that I’ve charted here is not the complete picture. It is the nature of indexical fields to shift over time and to be particular to individuals’ social experiences. Silverstein (1976)’s notion of indirect indexicality helps us see that we’re not working with a simple equation where a fast tempo equals a particular category or set of categories. Rather, fast tempo is associated with a field of qualities and stances that are then associated with categories. The consequence is that tempo reflects and constructs particular ideologies. Awareness of how certain types of people talk may actually make them talk that way and it may make someone perceive something that isn’t “really there”. These sorts of events happen daily and cause the associations to be replicated.

There does seem to be something at the heart of the fast speech indexical field, something from which the various indices are built. In examining the indexical fields of both fast speech and slow speech, I have been struck by the relationships to (a) emotion and (b) time, which I will turn to now.

Emotions and arousal

The literature on emotion is vast and there are a number of competing theories ranging from evolutionary to social constructionist.⁷ Those of us working on style usually have a social constructionist bent. One version of this for the theory of emotions is found in Averill (1980: 312): “an emotion is a transitory social role (a socially constituted syndrome) that includes an individual’s appraisal of the situation, and is interpreted as a passion rather than as an action.” Such appraisals necessarily involve categorization of

⁶ It is not completely clear to me where to put the “socially attractive” cluster of the bottom right (friendly, warm, polite, etc), though there are probably connections to be made between those characteristics and attempts to persuade (the car salesman acting like your friend).

⁷ Carl Lange even proposes that a racing heart isn’t a symptom of fear, it’s the cause of fear. The support for this comes from studying the emotions of people with injured spinal cords.

situations and any time we use the word “category” we ask: where did that come from? Moreover, identifying an emotion, whether as an experiencer or an observer, is also a form of categorization. All of these categories are based on beliefs, mediated through language, and ultimately brought about and maintained through social practices. With enough evidence, a person who displays anger can become an “angry person”. Likewise, as we observe nervous people chattering, we start connecting the speed of speech with the emotion.

Most of the work in the subfield of speech and emotion is focused on detecting emotion in speech or at least figuring out what acoustic cues listeners associate with different emotional states. The experimental work is not particularly theoretic, though it does usually ascribe to a belief in a discrete set of fundamental, basic emotions.⁸ The chart below, gives an example of some of the emotions that are most often described as basic.⁹

⁸ In recent years, psychologists have started talking about “the big six” and this is probably the dominant view in the field. But as Scherer (2003) points out, each researcher’s “big six” is slightly different.

⁹ Whether or not there are “basic” emotions, these lists are quite interesting for us to consider as we map out indexical fields. They offer both inspiration and an existing body of research. One of the best practices we should keep in mind is describing emotions as specifically as possible. “Anger” is more of a family of emotions than a singular feeling. “Hot anger” or “subdued anger” would make it easier to replicate results. See also Scherer (2003) and Eckman (1992).

Lazarus (1999a)	Ekman (1999)	Buck (1999)	Lewis and Haviland (1993)	Banse and Scherer (1996)	Cowie et al. (1999b)
Anger	Anger	Anger	Anger/hostility	Rage/hot anger Irritation/cold anger	Angry
Fright	Fear	Fear	Fear	Fear/terror	Afraid
Sadness	Sadness/distress	Sadness	Sadness	Sadness/dejection Grief/desperation	Sad
Anxiety		Anxiety	Anxiety	Worry/anxiety	Worried
Happiness	Sensory pleasure	Happiness	Happiness	Happiness Elation (joy)	Happy
	Amusement Satisfaction Contentment		Humour		Amused Pleased Content Interested
	Excitement	Interested Curious Surprised			Excited
		Bored		Boredom/indifference	Bored Relaxed
		Burnt out			
Disgust	Disgust Contempt	Disgust Scorn	Disgust	Disgust Contempt/scorn	
Pride	Pride	Pride Arrogance	Pride		
Jealousy		Jealousy			
Envy		Envy			
Shame	Shame	Shame	Shame	Shame/guilt	
Guilt	Guilt Embarrassment	Guilt	Guilt Embarrassment		
Relief	Relief				Disappointed
Hope					
Gratitude					Confident
Love			Love		Loving Affectionate
Compassion		Pity Moral rapture Moral indignation			
Aesthetic					

Figure 2: Cowie and Cornelius (2003) review recent lists of "key" (basic) emotions

Regardless of the number of emotions that researchers are describing, there is a strong tendency to plot the emotions against specific dimensions. The most typical axes are **valence**, which really means “pleasant vs. unpleasant” and **activity**, which sets up a contrast between active emotions (joy) and passive ones (sadness).¹⁰

¹⁰ Other dimensions that have been proposed have to do with **power, control, attention, competence, locus of causation, potency, and dominance**. Potency is probably the most common “third dimension” (researchers like to keep the number of dimensions down to two or three). All of these proposals are interesting from ideological standpoints, though I will be focusing on activity. For a little more detail about various emotional dimensions, see Scherer (2003: 244).

I have mentioned Scherer several times already—he is probably the preeminent researcher in the field of speech and emotion and his (2003) review is the most comprehensive introduction.¹¹ As he mentions, the majority of studies have looked at F₀, energy, and speech rate because those have been the easiest parameters to measure. I’ll discuss speech rate in a moment, but first a few general observations: fear, joy, and anger have high mean F₀ while sadness has a lower mean F₀. The “active” emotions also have more F₀ variability and vocal intensity.¹²

This notion of activity (also known as **arousal**) is crucial for us to understand. I believe it is a core part of how indexical fields for tempo get constructed. Activity is, of course, measurable in terms of heart-rate, respiration, glucose uptake, epinephrine release, etc. When we look at the indexical fields for different tempos, we see that the active emotions (rage, joy) are associated with fast talk while the passive emotions (sadness, boredom) are associated with slow talk.

	Stress	Anger/rage	Fear/panic	Sadness	Joy/elation	Boredom
Intensity	↗	↗	↗	↘	↗	
F0 floor/mean	↗	↗	↗	↘	↗	
F0 variability		↗		↘	↗	↘
F0 range		↗	↗(↘)	↘	↗	↘
Sentence contours		↘		↘		
High frequency energy		↗	↗	↘	(↗)	
Speech and articulation rate		↗	↗	↘	(↗)	↘

Figure 3: Acoustic patterning of basic emotions (Scherer 2003, based on Johnstone and Scherer 2000).

These results are generally consistent across studies, with few exceptions. One of the most interesting is to consider what happens when we split up men and women.

	Pitch				Intensity		Timing	
	Mean	Range	Variance	Contour	Mean	Range	Speech rate	Transmission duration
Anger	>>	>	>>		>> _M , > _F	>	< _M , > _F	<
Disgust	<	> _M , < _F			<		<< _M , < _F	<
Fear	>>	>		↗	=>			<
Joy	>	>	>	↘	>	>		<
Sadness	<	<	<	↗	<	<	> _M , < _F	>

Explanation of symbols: >: increases, <: decreases, =: no change from neutral, ↗: inclines, ↘: declines. Double symbols indicate a change of increased predicted strength. The subscripts refer to gender information: M stands for males and F stands for females.

Figure 4: Relationships between emotions and acoustics, from Ververidis and Kotropoulos (2006: 1171).

¹¹ Scherer’s great strength is laying out the literature. His personal model of how emotions are produced and perceived in speech is a little hard to follow in the (2003) review, though it does have a “probabilistic functionalism” that focuses on the environment and builds in uncertainty. Like other researchers, he allows for sociocultural factors, but he isn’t particularly interested in those. (This is a little bit curious since he is interested in the idea that emotions are inherently variable and depend upon the appraisal of a situation.) Mostly, he is looking for universal expression of emotion in speech that are rooted in psychobiology. A lot more research has been done along universal expressions of emotion in the face—especially by Ekman. If you are familiar with that name, then you can consider Scherer the “Ekman of the voice”, as others have put it.

¹² Bachorowski (1999) reports that the difference among active emotions like joy and anger may be about the contours of F₀—it may decrease over time for anger, but increase for joy.

“The same emotion in man and woman are nonetheless dissimilar in tempo,” said Nietzsche (1886) and while the Scherer (2003) pairing of anger pairing and fast tempo is the norm, others find a gendered response even when the situations are similar: the men in these studies speak with slow (“cold”, “subdued”) anger, while the women speak quickly (Heuft et al 1996, Iida et al 2000, also cited in Ververidis and Kotropoulos 2006). The relationship between anger and gender is ideologically in its nature and in its consequences: there is a sense in which hot anger is about a lack of control. Losing control, of course, is unmanly, though it is long-standing part of female gender performance. There’s quite a lot to say about gender and tempo and I’ll have a bit more to later on (“Case 3: A woman inside Captain Kirk”). I can’t hope to be exhaustive, though.

One of the main associations with speed is change—if you speed up a video recording, changes become all the more noticeable and dramatic. Change, in turn, can point to volatility. As can any of the active emotions, since it is in their nature to sweep in and out of our lives. The passive emotions creep.

We can describe these processes as instances of iconization. Iconization turns linguistic differences into representations of the social contrasts they index, as if the linguistic variable somehow showed the group’s true essence (Irvine 2001: 33). Social identity is attached to linguistic conduct, including tempo. By this process, swiftness in speech is taken to mean speed in the mind or variability in emotion; slowness of speech shows the slowness of the body or the torpor of the soul.

Time ideology

In the previous section, I argued that arousal/activity was one of the main themes running through the indexical fields for tempo. In this section, I will briefly sketch out another theme: the relationship to time.

It is obvious that speaking quickly allows you to either (a) fit more into a given time period, or (b) take less time than someone at a moderate or slow pace. There are ideological consequences for how much time one takes and how much content one gives. But these can go in a couple directions. On the one hand, a person speaking quickly may communicate that they don’t want to waste your time and/or they don’t want you to waste theirs. But speech rate may also indicate power relationships, along the lines of Michael Caine’s thoughts: “The basic rule of human nature is that powerful people speak slowly and subservient people quickly—because if they don’t speak fast nobody will listen to them” (1992).¹³

Speech unfolds over time, so the rate of that unfolding is tightly associated with time itself. This is an iconic relationship: how much time you use is taken to reveal what you think about time. It is “natural” to see fast talkers as “in a rush” because they can theoretically get more accomplished if they whiz through conversations. But like other icons, tempo is “mediated through social convention and subject to the historically specific interpretative habits of its users” (Mannheim 2000: 107).

The meaning of speech rate emerges from interactions, and each individual interaction helps build up the indexical field. There are a number of studies that have linked judgments of competence to speech rate.¹⁴ These studies find that speakers who use a faster tempo are rated as more competent than those with a slower tempo. Street (1985) adds a wrinkle. Subjects listened to student interviews and rated the

¹³ And this is, in fact, the relationship that Irvine (2001: 36) reports for Wolof nobles (slow tempo) and griots (fast tempo).

¹⁴ For example, Smith et al (1975).

competence of both interviewers and interviewees. Like previous studies, the faster talkers were seen as more competent. But Street also had the competency ratings as done by the interviewers and interviewees themselves. He found no correlation between outside observers and internal interview participants. In fact, ratings sometimes went in opposite directions, with interviewers giving high competency marks to interviewees who had longer response latencies and slower speech rates.

In real life, people encounter problems of various levels of difficulty. Those that are complex are seen as taking longer than the simple ones. This is mapped on to internal thinking and language processing, too: complex thoughts take longer to think and articulate. By extension, someone who is capable of talking quickly may be “thinking quickly”. That is, they have more intellectual horsepower and operate at a faster pace. On the opposite side, those who speak slowly in normal conversation are obviously taxed by the cognitive load of simple things.¹⁵

The contrarians in Street’s study that gave higher competency ratings to partners who spoke more slowly may have been judging a different sort of competency. External observers were almost certainly looking for competency as a reflection of the individuals. When two people are conversing, however, they must take turns speaking, which makes them focus on their partner’s speech as well as their own (Clark 2002). A partner in a conversation is swayed to consider such interactional, social competency. Someone speaking quickly may not be adequately considering their interlocutor. That can be a misjudgment of the situation and a sign of emotional unintelligence. It *is* possible to get external raters to appreciate the interactional aspects: if they are told that speakers are explaining a difficult topic to an audience of beginners, the incompetency bias for slow speech disappears (Brown et al 1985).

This interactional aspect of tempo extends to how speed is deployed. New Yorkers can’t just talk quickly—even with other New Yorkers. There are requirements to be a good conversational partner, as anthropologist Michael Agar (1994) points out:

New Yorkers speak more quickly and shift topics more rapidly. We all know that. But the way it’s done is fascinating. At least it fascinated me when I first moved to New York. Someone has the floor and talks. As soon as I know what they are going to say, I can jump in, finish the sentence to show I understand, and take off into my own turn. The northern California I know isn’t like that. Someone talks, and I lie back and listen and let them roll for a while. When they’re done, there’ll be a pause that will flash like a green light to announce that someone else can have the floor.

Agar’s statement would lead us to conclude that there are sanctions—even in New York—for the person who jumps in without knowing how the previous speaker is going to finish the sentence. Without looking at the interaction of speakers, we can’t resolve what meaning will come to the surface. Is the interruption productive or offensive? Analyzing tempo requires an interactional mindset, too.

Context, interaction, performance, Captain Kirk

Until now this paper has focused on sketching the indexical field of tempo. What characteristics are associated with “fast talk”, for example, and what sort of connections can we find between them? I’ve suggested that the indexical field for tempo reiterates many of the ideological connections we see

¹⁵ A number of researchers have found that faster speakers are more persuasive speakers, in tracing what this means, Miller et al (1976) found that faster speakers were seen as more knowledgeable, intelligent, and objective.

elsewhere: between cognitive abilities and language, between gender and dominance. In this section, I turn my attention to variability in tempo and perform a close reading of several performances, which I think can teach us about how tempo is used stylistically. These also ground us in particular contexts and give me the opportunity to stress how important interaction is in generating meaning.

On acting

Actors are a good source for studies of style since they make vivid the cues that are more mixed and grey in real life.¹⁶ “The act that one does, the act that one performs, is, in a sense, an act that has been going on before one arrived on the scene” (Butler 1988: 526). Butler is talking about gender, but this idea applies to acting as well. Actors don’t really create anything out of whole cloth. They assemble bits and pieces. It would be difficult to analyze the acoustic signal of wooden acting, since we’d be measuring perceptions of lack, but even histrionic, scene-chewing samples offer us speech cues associated with various social categories. Again, the assumption is that actors use stylistic resources that their audiences can be expected to understand. If audiences uniformly agree on what a performance expresses, it doesn’t necessarily matter what the intention was. We’re after that shared social meaning and the components that comprise it, though we may be giving up the psychophysiological effects on the voice that happen under natural conditions.

The bulk of the research on emotion and speech has used professional or lay actors.¹⁷ The field relies on the fact that if vocal portrayals are reliably recognized by judges, they must reflect real-life expression patterns at least to some extent.

There can be little doubt that simulated vocal portrayal of emotions yields much more intense, prototypical expressions than are found for induced states or even natural emotions...[And] it cannot be excluded that actors over-emphasize relatively obvious cues and miss more subtle ones that might appear in natural expression of emotion. (Scherer 2003: 232, see also Scherer 1986: 144)

This seems to be more of a problem for those who are interested in the detection of underlying states. However, I am attempting to understand sociocultural norms, so I can take greater comfort in shared recognizability and the idea that any observable expression of emotion is a *performance* of the emotion, even if it is naturally occurring. Still, work in this direction should be balanced by ethnographies involving natural speech.

It’s important for us to keep performance in mind as we look at styles. If a characteristic can be detected, it is doing something, not merely revealing an internal state. Which is to say, stylistic signals are like speech acts—they do something once they are released into the world, whether intended or not. Tempo doesn’t usually annul a marriage or grant land rights, but it does shift an audience’s perspective about the nature of the speaker and the interaction. Psychologists seem to miss this or at least turn away from it. Speech is not just expressive. If I send off cues that I am “emotionally out of control”, your interaction

¹⁶ Writers create scenes of dramatic interest, so that there is also a higher proportion of arousal in a scene than in daily life.

¹⁷ There has been some use of real-life recordings (pilots in dangerous flight conditions, reporters at emotional news scenes, therapy sessions, talk shows and game shows), but since researchers claim to be interested in detecting expressions of emotion, they are usually left assuming what emotion the people were feeling. As of Scherer (2003), no one had actually tried to match real-life recordings with acted performances. Such a study would seem to be important for understanding the limits of using actors in emotion research.

with me shifts; even an outsider's perception of me shifts. Moreover, whatever my underlying state and intention, it can certainly be useful to be *seen* to be out of control.

In the sections that follow, I emphasize the importance of context and interaction in understanding style. Research in speech and emotion usually involves giving actors some standard sentences and asking them to read these sentences in various emotions. This enables researchers to control for content, but it gives them false confidence that they are avoiding interference with semantic and syntactic cues. I'd like to suggest that listeners infer emotional meaning (or social meaning since my concerns are not solely with emotion) based on an interaction of linguistic structure, context, and vocal cues.¹⁸ In between natural recordings and elicited acting are scenes from plays, films, and television shows. These scenes ground actors in particular contexts, with particular interlocutors. The actors are still relying on sociocultural norms, but we gain information by looking at situated portrayals, even if we do lose control over the content. I suggest not that studying YouTube videos is *better* than previous methods, but that it is an important addition.

Starburst: The William Shatner story

William Shatner has a distinctive style. Here is one amateur impersonator, revealing the secret of “doing William Shatner”:

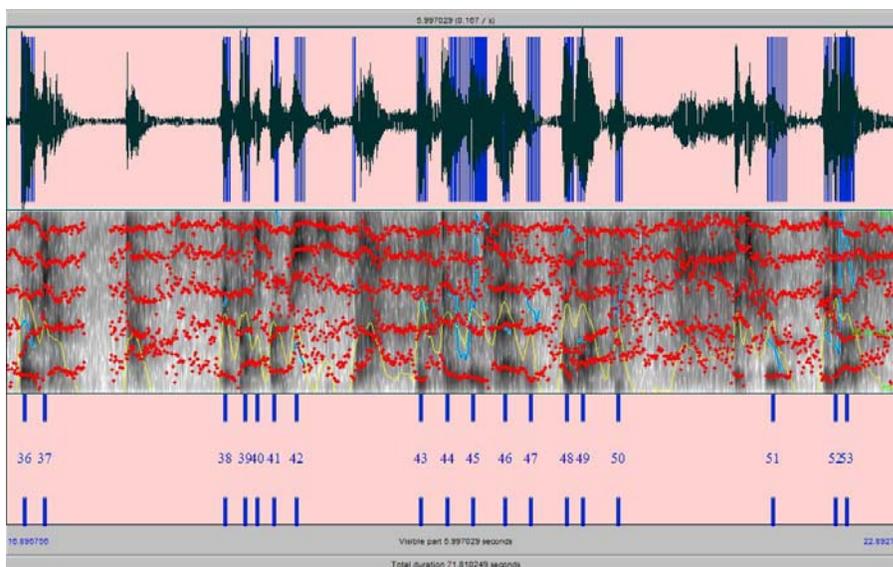


Figure 5: "William Shatner talks in fast bursts and moves his head a lot it's true he does", http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kids_impersonation_of_Shatner_converted.wav

The key, as this young man reveals, is inserting pauses that are followed by rapid bursts. In looking at impersonators, I'm struck by how similar they are to William Shatner when he is in bursty-mode. The main exaggeration is that impersonators *only* do bursts, while William Shatner deploys bursty speech (a little) more selectively.

One of the questions we want to answer is what this burst-style is about. To get to that answer, we must consider that whatever role an actor is performing, they are also performing as an actor. Even the strictest

¹⁸ See also Ladd, Silverman & Scherer (1983) or Scherer (1986: 145) for a very brief linguistic critique of psychologists' assumptions.

Stanislavskian is crafting a performance.¹⁹ Recognizing the social role of the actor is, in fact, central to understanding what William Shatner is doing.

Shatner's pause-and-fast-burst approach enacts a kind of **drama** on the phrasal level. Here, you have an actor who deploys bursty tempo during scenes that are the most dramatic. These bursts of drama are especially elevated because the other actors in the scene are likely to be much more regular in their tempo. For example, Leonard Nimoy plays Mr. Spock. Mr. Spock is a Vulcan, a race that has turned away from emotion—his speech is decidedly even.

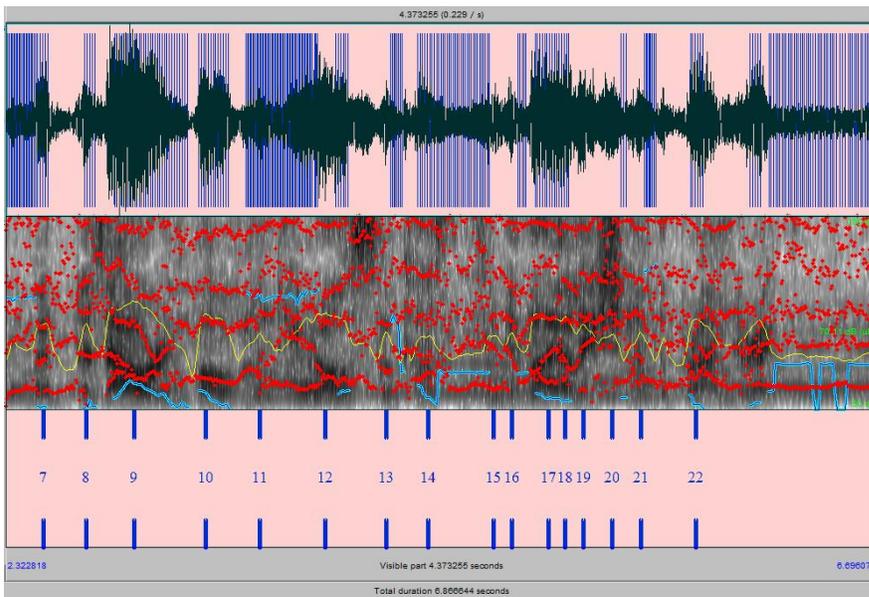


Figure 6: Leonard Nimoy's Mr. Spock speaks in a regular tempo (somewhat swiftly--4.81 syll/sec). http://www.stanford.edu/~tylers/notes/socioling/Sounds/Spock_Our_involvement.wav

Spock speaks at a fairly fast clip, which is in keeping with our expectations about the social categories indexed by fast speech—Spock is supposed to be inhumanly intelligent, knowledgeable, and objective (in the show, he can't use the word "logical" often enough). Spock's rate of speech matches his ability to process complex information quickly, demonstrating the iconic relationship among time-of-thought, effort-of-thought, and rate-of-speech.

It's reasonable to ask whether there's a conflict between Spock's emotionlessness and the fact that fast speech is also associated with "active emotions". I believe that the right way to think about fast speech is that it indexes "activity" more fundamentally than "active emotion." Spock is certainly an active character—no one is more "on" than he is, though that alertness isn't supposed to correspond to a heightened emotional state. It's also worth noting that stylistic variables may well index contradictory characteristics and figuring out how these are resolved is a worthy goal for future research. In this case, it isn't hard to imagine the calculus that settles the conflict. The alternatives to a fast-talking Spock would be an average-speed Spock or a slow-talking Spock. First, the connection between slowness of speech and slowness of mind seems to trump the association of fast speech and emotion.²⁰ But not only does the

¹⁹ Non-actors also offer us performances, but the awareness of the performance is a little harder to assess.

²⁰ In my pilot survey, mental slowness is among the most frequent associations with slow speech. But people do allow that teachers explaining something speak slowly, too. And Mr. Spock does seem to slow down when it's necessary to get something across to the simpleton humans. He does not *only* speak quickly.

slow speech indexical field repel the character, the indexical field of fast tempo probably exerts an attractive force. Fast speech is, by definition, marked compared to the average, and such markedness is appealing to an actor, especially if other cues can stave off unwanted associations. Indeed, Spock does spend an awful lot of time asserting his emotionlessness.

Kirk, by contrast, is meant to have scrappier street smarts and to be a man of passion. He is often emotionally aroused—and he is also the captain, giving him the luxury of long pauses that Michael Caine’s butlers would have to wait for. These two factors could motivate the combination of fast speech and pauses that look like burstiness, but character construction is not the only thing going on.

A central part of what’s happening with Shatner is that he is using tempo to activate the notion of drama. That is, the aroused emotions are the ones that suit Kirk’s character (impetuous leader) and they suit Shatner (the lead actor, a standout). There is an old design adage that says, “First, draw a distinction”. Shatner has taken this to heart and he’s come up with a unique verbal style that sets him apart from other actors. Tempo is a crucial part of this.

Case 1: Variation in a monologue

In the scenes I analyze, I focus on what tempo seems to be expressing and how it interacts with the various situations that Shatner and Captain Kirk find themselves in. The second and third case studies are interactional, but *Star Trek* often gave Shatner concluding monologues and those are interesting to look at, too.

In the episode, “The Omega Glory” the Enterprise comes upon a planet in which the Yongs (Yankees) and the Kohms (Communists) have fought wars that Earth avoided. Kirk realizes that they have been fighting based on a document he knows well—only the Yongs have completely misinterpreted it.

The climax in Figure 7 is about 4.06 syll/sec, though as you can tell by looking at the chart, it is somewhat bursty. Shatner starts slow within it, pauses, and then speeds up.

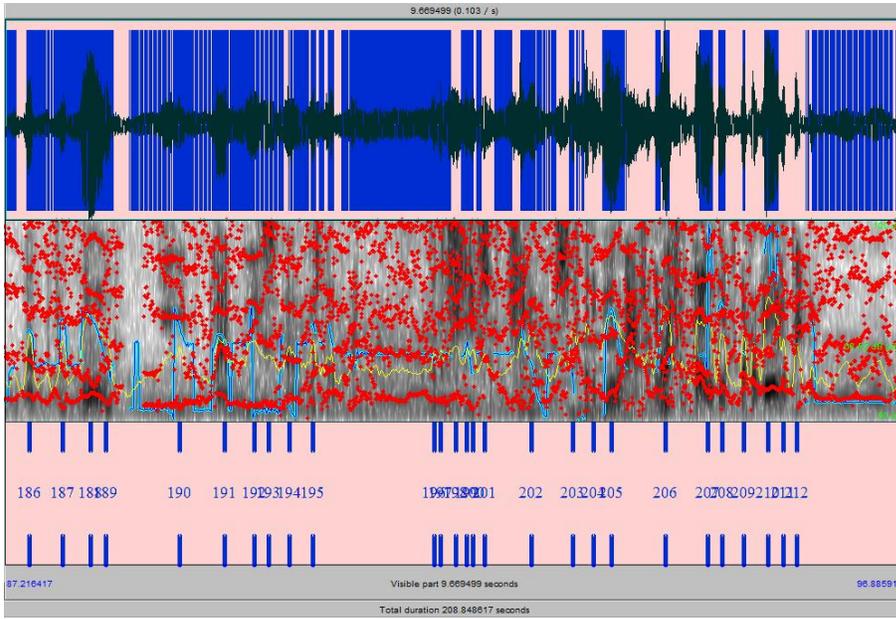


Figure 7: Kirk talking about the Constitution: “What you call, ‘Ee’d Plebnista’ was not written for the chiefs or the kings or the warriors or the rich or the powerful, but for all the people!”
 (See <http://www.youtube.com/watch?v=gutY7NnNuyA> at 0:56)

But not only does tempo vary within an utterance, it also varies across utterances. Here is the complete monologue, following it I show the speech rates graphically.

- (1) This was not written for chiefs. (**4.035 syll/sec**)
- (2) (*Grumbling*)
- (3) Hear me! Hear this! (**2.618 syll/sec**)
- (4) (*Pause*)
- (5) Among my people (**5.208 syll/sec**)
- (6) (*Pause*)
- (7) we carry many such words as this from many lands (**4.237 syll/sec**)
- (8) (*Pause*)
- (9) many worlds (**4.148 syll/sec**)
- (10) (*Pause*)
- (11) Many are equally good and are as well respected, but wherever we have gone (**4.985 syll/sec**)
- (12) (*Pause*)
- (13) no words have said this thing of importance (**2.385 syll/sec**)
- (14) (*Pause*)
- (15) in quite this way (**2.452 syll/sec**)
- (16) (*Pause*)
- (17) Look at these three words written larger than the rest, with a special pride never written before (**4.080 syll/sec, though this phrasing builds up speed even in itself**)
- (18) (*Pause*)
- (19) or since. Tall words proudly saying, “We the People.” (**2.201 syll/sec, also slow, fast, slow internally**)
- (20) (*Very short pause*)

- (21) That which you call Ee'd Plebnista was not written for the chiefs or the kings or the warriors or the rich or the powerful, but for all the people! (**4.06 syll/sec**)
- (22) (*Pause*)
- (23) Down the centuries, you have slurred the meaning (**3.144 syll/sec**)
- (24) (*Tiny pause*)
- (25) of the words (**2.201 syll/sec**)
- (26) (*Tiny pause*)
- (27) We (**drawn out for 1.156 seconds**)
- (28) (*Tiny pause*)
- (29) the people of the United States (**3.823 syll/sec**)
- (30) (*Pause*)
- (31) in order to form a more perfect union, establish justice, ensure domestic tranquility, provide for the common defense (**4.439 syll/sec**)
- (32) (*Pause*)
- (33) promote the general welfare, and secure the blessings of liberty (**3.829 syll/sec**)
- (34) (*Pause*)
- (35) to ourselves and our posterity (**2.945 syll/sec**)
- (36) (*Pause*)
- (37) do ordain and establish this constitution.” These words (**2.987 syll/sec**)
- (38) (*Pause*)
- (39) and the words that follow (**4.211 syll/sec**)
- (40) (*Pause*)
- (41) were not written only for the Yangs, but for the Kohms as well! (**3.409 syll/sec**)

The whole thing takes over a minute and a half, which is an eternity in television time. Tempo variation gives Shatner the ability to hold on to attention by animating the words. When Shatner says (23), he draws out “slurred” and “meaning”—he is not only being corrective in his proclamation, but his lengthening is the anti-mimetic representation of slurring.²¹ Later, the preamble to the Constitution is read slowly and significantly, but as it goes on Shatner picks up speed. The speech can’t be completely ponderous and it would probably be dull if he labored over all the lines about forming, establishing, ensuring, etc. Those phrases are beside the point, anyhow. The goal is to say that over time we sometimes forget to align our actions with our beliefs.

Variation in speech tempo can help keep up the listener’s attention. The juxtaposition of fast and slow allows Shatner to stress some ideas and rush over others. He is also attempting to give a unique reading of this monologue, something that another actor wouldn’t or couldn’t do. Thus the variation in tempo dramatizes the text and reflects Shatner’s role as an actor. It’s through choices like this that the “William Shatner” style emerges: theatrical, bursty, unusual.

²¹ Drawing these words out is also for the virtue of his 20th century audience, since the message is meant for them, too.

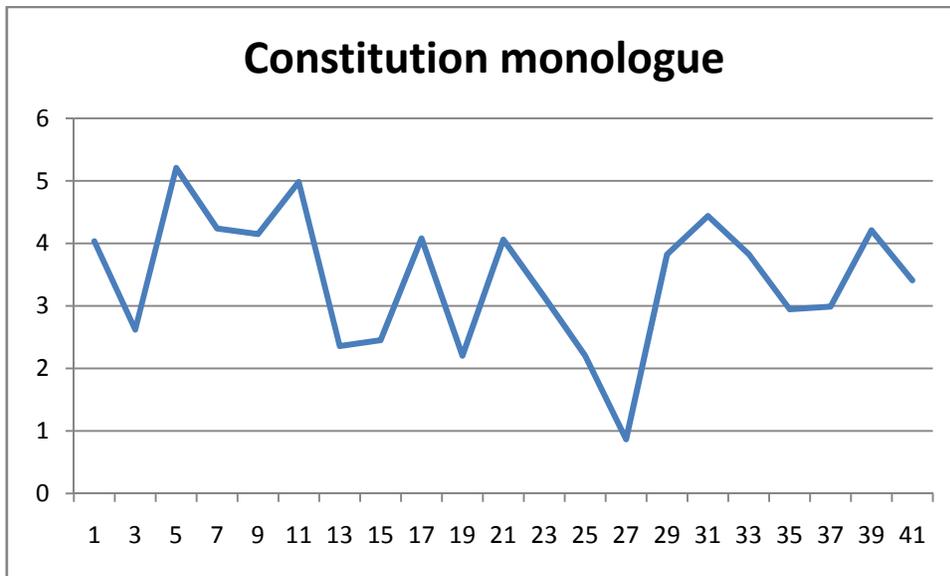


Figure 8: Kirk's monologue about the Constitution. The average tempo is 3.454 syll/sec. Most of the sections are within one SD of that average, but sections 13, 19, 25, and 27 are slower and sections 5 and 11 are faster.

Case 2: Superburst

An episode synopsis won't be necessary for the scene below. The basic context is that the ship's doctor (Leonard "Bones" McCoy) is concerned with the health of Mr. Spock and wants Captain Kirk to turn the ship around and go to Spock's home planet, Vulcan.

This scene demonstrates tempo in interaction. And it also contains one of Shatner's most bursty exclamations. You'll be able to hear it in the recording and see it in Figure 14. But I also have been trying to come up with a way to measure and report degrees of burstiness.

We can draw a parallel to network traffic engineering, which attempts to measure the burstiness of packets traveling across the Internet (such burstiness affects the quality of some applications, like voice and video). Of the several measurements I have tried, the one that seems to work best is *variance/mean*. That is, I input how much time there is between syllables and calculate the variance (the dispersion of the data). Then I divide the variance by the mean number of syllables in the utterance. This provides a measure of how spread out the data is; the bigger the ratio, the more it is characterized by clusters—what we'd call bursts.

At the start of our scene, Dr. McCoy (DeForest Kelley) is urging Captain Kirk (William Shatner) to get Spock to his home planet, Vulcan.

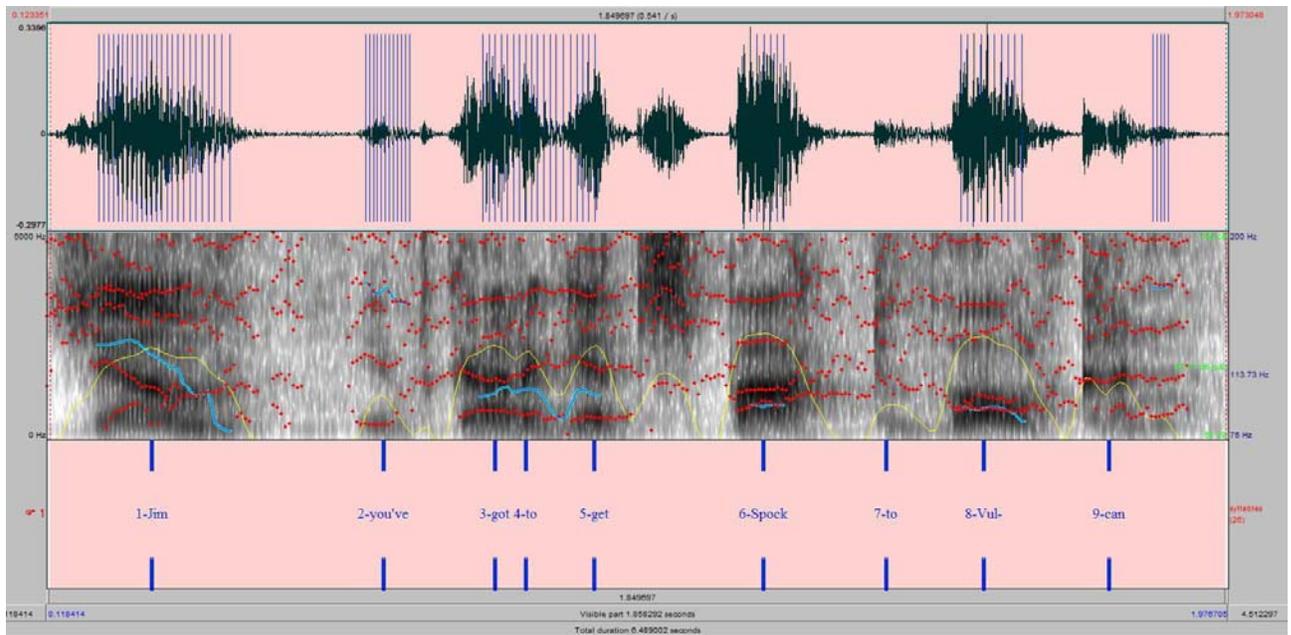


Figure 9: Bones, “Jim, you’ve got to get Spock to Vulcan”; 4.552 syll/sec. http://www.stanford.edu/~tvlers/notes/socioling/Sounds/Bones_Kirk_1_dialog.wav

This is a 1.977 second clip, but he squeezes in 9 syllables (4.552 syll/sec). In the clip, we can hear the term of address—Jim!—get some space, followed by a rush of words, with “Spock” as the one that stands out most. Following the apostrophical “Jim!”, the speech is fairly even. The overall burstiness score is 0.00182.

Here’s how Captain Kirk responds. He also calls his friend by name and then starts off along a fairly regular tempo. The speech rate here is 4.893 syll/sec and the burstiness is 0.00329.

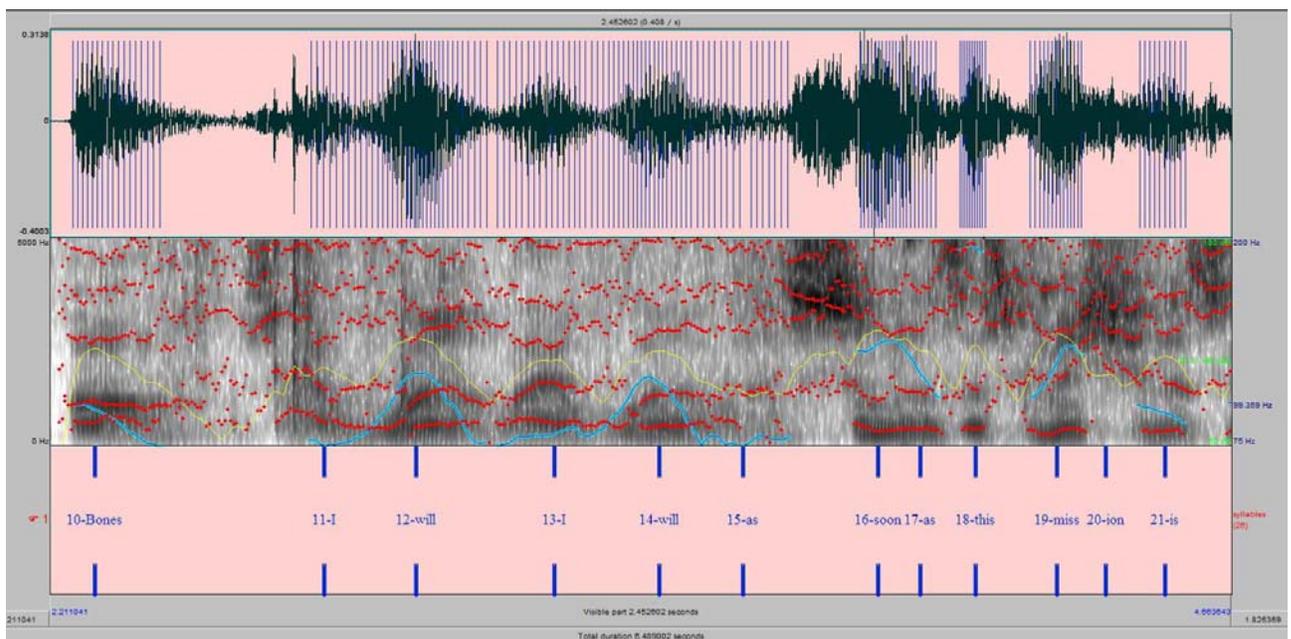


Figure 10: Kirk, “Bones, I will, I will, as soon as this mission is—”, http://www.stanford.edu/~tvlers/notes/socioling/Sounds/Bones_Kirk_1_dialog.wav

Bones doesn't want to be placated. His message is urgent. He interrupts Kirk, shouting "No!" before Kirk can say "completed". If we use the whole time span here to calculate a speech rate, we'll get something very slow like 2.892 since "No!" has so much space after it. If we restrict ourselves to the "Now, right away" portion, then the rate jumps up to 4.735 syll/sec. The burstiness score of the whole clip is 0.0223.

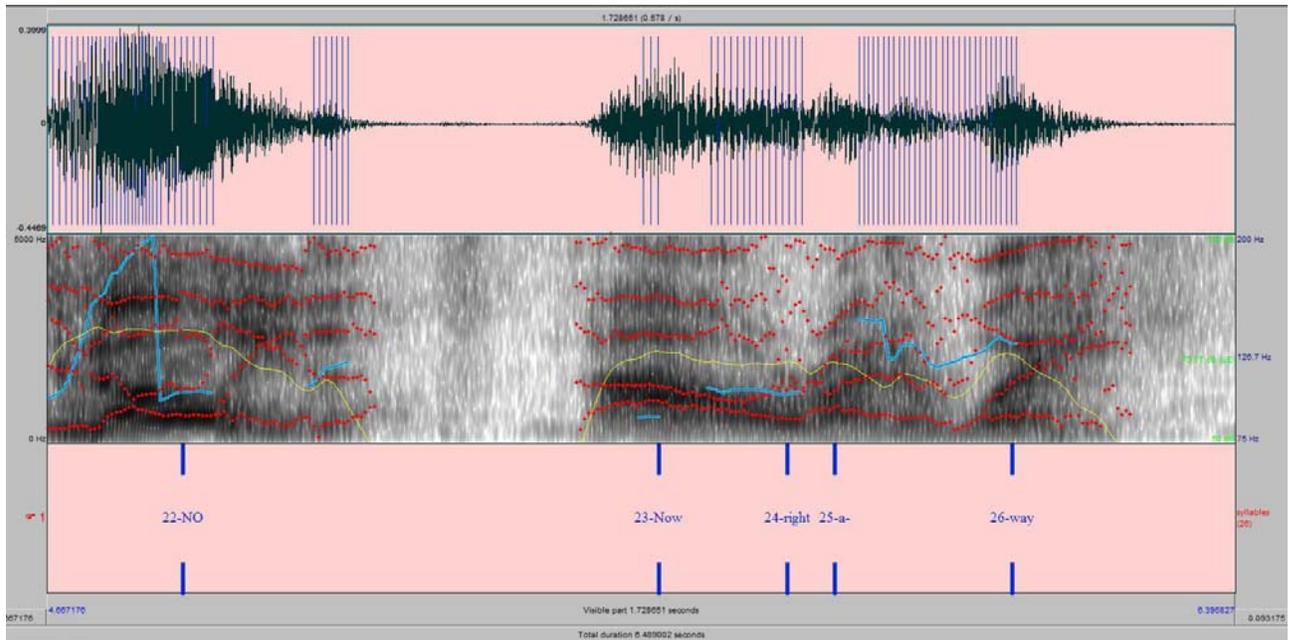


Figure 11: Bones, "No! Now. Right away.", http://www.stanford.edu/~tylers/notes/socioling/Sounds/Bones_Kirk_1_dialog.wav

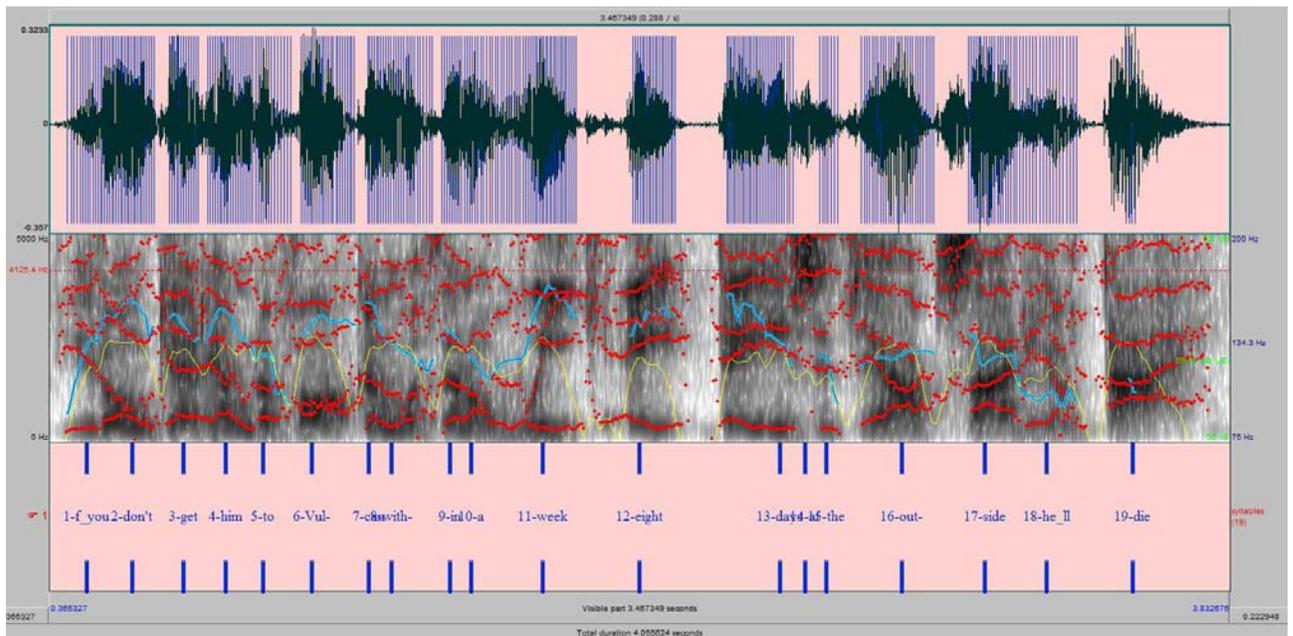


Figure 12: Bones, "If you don't get him to Vulcan within a week, eight days at the outside, he'll die." http://www.stanford.edu/~tylers/notes/socioling/Sounds/Bones_Kirk_2_if_you_dont.wav

Bones continues with the consequences—he’s speaking rapidly now (4.957 syll/sec) since he is anxious. The speech is even (0.00133 for burstiness), which is probably what makes this sound more worried-and-I’m-trying-to-persuade-you than out-and-out panicked.

Then, in an intense whisper, Bones repeats, “He’ll die, Jim”. The rate (3.229 syll/sec) is what we’d call average, but it sounds rather slower since everything else has been faster. It’s pretty even (0.00672 for burstiness).

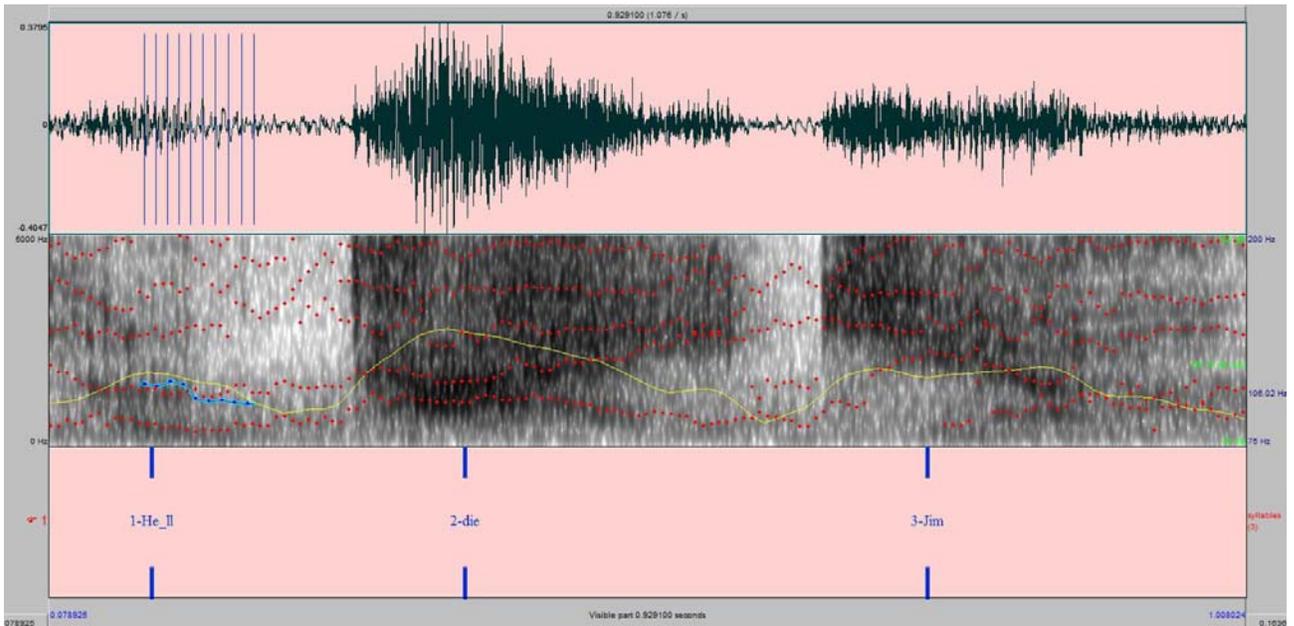


Figure 13: Bones, “He’ll die, Jim”, http://www.stanford.edu/~tylers/notes/socioling/Sounds/Bones_Kirk_3_he_ll_die_Jim.wav

This switch down from speed and into whisper is important for both the character and the actor. It takes advantage of other stylistic cues—the urgency of an intense whisper, for example. We can’t isolate tempo from other acoustic signals and it’s often the combination that give meaning. As we’ve seen with Shatner, the variation in tempo also holds up interest—the key to music, many people believe, is variation within a pattern. That aesthetic principal governs actors, too.

But let’s see how Shatner/Kirk responds since it is practically a Kevin Pollack impersonation.

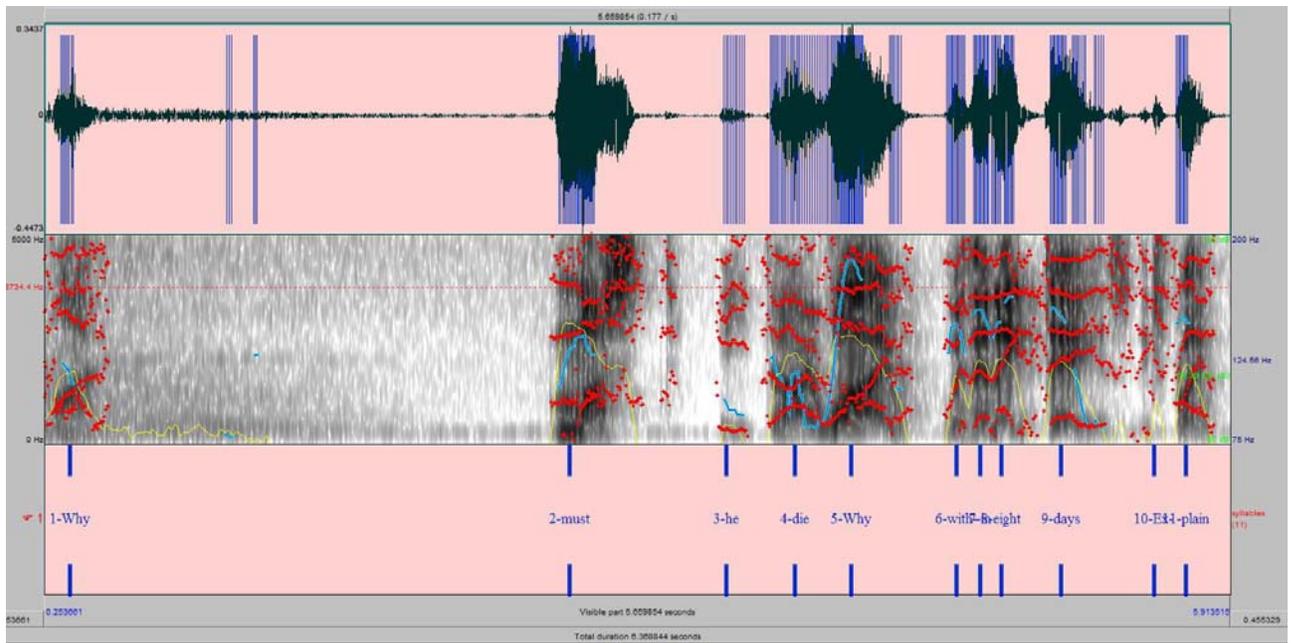


Figure 14: Kirk, “Why...must he die? Why...within eight days? Explain.” http://www.stanford.edu/~tylers/notes/socioling/Sounds/Bones_Kirk_4_Shatner_super_burstv.wav

If we calculate syll/sec over this whole time, then it’s really slow (1.944 syll/sec) because Shatner puts such a long pause between “Why” and “must”. Indeed, it sounds at first that the question is simply “why?” The continuation is unexpected. The tremendous pause does interrupt a phrase, but it does so in such a curious way, that I propose we should also look at the rate of the rest, which turns out to be a more normal 3.078 syll/sec rate. But notice the burstiness—there’s a huge pause after the first “Why” and another sizeable one after “must”. There’s a pause after the second “Why”, too, but the lack of one before it is strange to the ear. The concluding “Explain” is said at a break-neck speed (5.133 syll/sec).

So far our burstiness values have been small and haven’t varied a great deal. At this point, however, burstiness jumps up to 0.0762. That’s still a decimal, so let me show it in a chart:

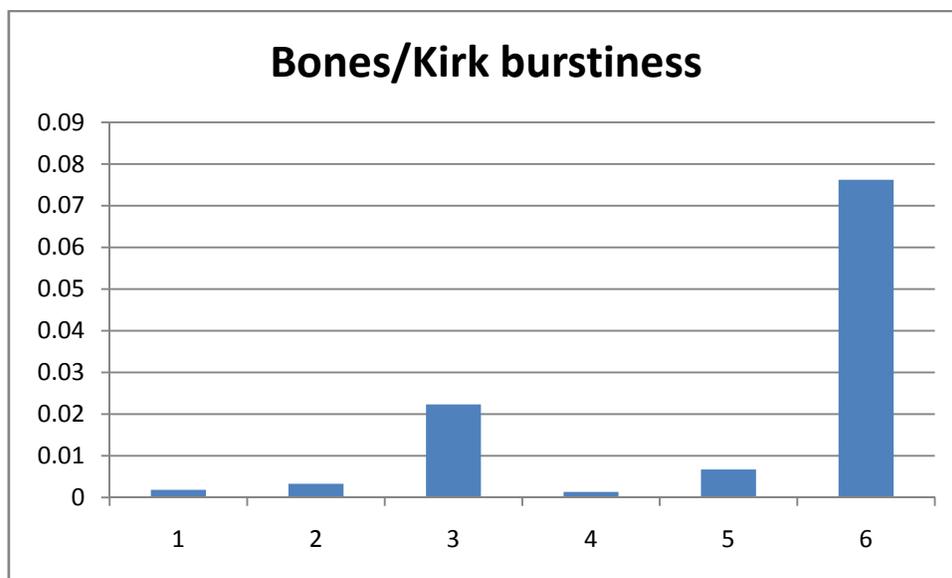


Figure 15: The spikes in burstiness correspond to the burstiest sounding moments

Kirk’s final utterance is over four times burstier than the scene average. I don’t have space (or time) to prove that my measurement of *variance/mean* is the best measure of burstiness, but the results here look promising since they align with what we hear with our ears and see with our eyes. One of my claims is that Shatner sets himself apart by using the bursty strategy. The second-most bursty utterance is Kelley’s but it pales in comparison and all of its burstiness really comes from the pause after an apostrophic address to Kirk. Such addresses are routinely followed by pauses, so there is nothing particularly unusual in his usage.

In Shatner’s first turn (“Bones, I will, I will...”), Kirk is trying to reassure and pacify, so burstiness isn’t really appropriate. But in any elevate circumstance, the burstiness strategy does suit Kirk since Kirk is a highly aroused character. Burstiness also gives a reading of the script that is unique and which distinguishes him from the actors both on the *Star Trek* set and elsewhere. I won’t present examples in this paper, but Shatner regularly deploys burstiness, whether playing an alcoholic ex-priest, a distressed veterinarian, or a racist Southern agitator. It’s how he distinguishes himself.

Case 3: A woman inside Captain Kirk

I turn now to the curious case of Dr. Janice Lester, an old flame of Captain Kirk who could’ve been a starship captain but for the sexism of Starfleet.²² The combination of the broken romance and crushed ambition twisted Lester and when she got a hold of a “life-energy transfer device” among the ruins of an ancient civilization, she trapped Kirk and switched bodies with him.

What this meant, of course, is that William Shatner got to play a woman pretending to be Captain Kirk.²³ In the scenes that follow, Lester-in-Kirk’s-body is trying to get to Benecia colony. She’s plotting to execute the senior officers for mutiny (corporal punishment is illegal in Starfleet, but she’s going to do anyhow). They’ll be buried on Benecia.

²² Yes, even in the year 2269.

²³ And actress Sandra Smith was left playing Captain Kirk in Janice Lester’s body. Smith is unconvincing in these scenes—her Kirk is dreamy and feminine.

Immediately after ordering the executions, Lester goes to the bridge and begins to give orders. Her speech here is a remarkable 7.485 syll/sec. This is an extension of the extreme emotional state we've seen in the mutiny trial moments before. Regarding "Kirk's" courtroom behavior, Dr. McCoy has said, "I've seen the captain feverish, sick, drunk, delirious, terrified, overjoyed, boiling mad. But up to now, I have never seen him red-faced with hysteria." We are meant to understand—and it's hard to miss—that the behavior of "Kirk" is not only erratic, it's unlike any other emotion he ever displays.

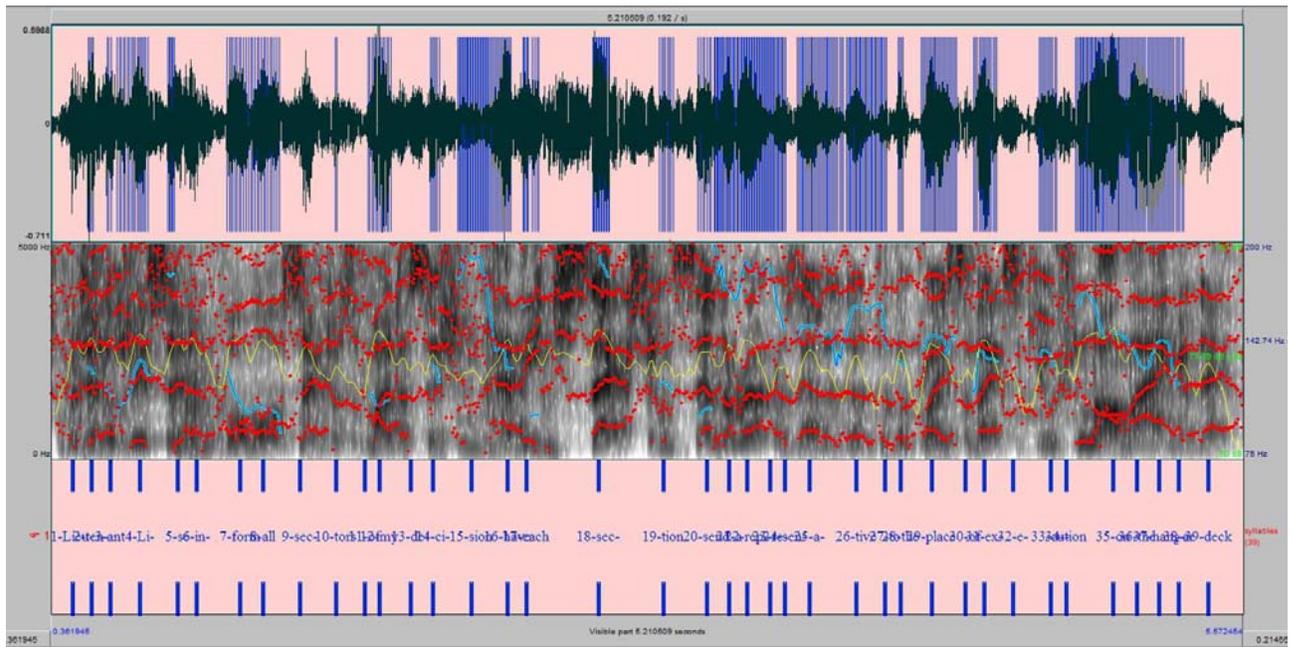


Figure 16: Shatner performs as a woman inhabiting Captain Kirk's body: "Lt. Lisa inform all sectors of my decision. Have each section send a representative to the place of execution on the hangar deck." 7.485 syll/sec. <http://www.stanford.edu/~tvlers/notes/socioling/Sounds/Kirk as woman bridge 1a.wav>

Right after this utterance, Lester asks the helmsman, Chekhov, how long it will take to get to Benecia colony. This utterance is even faster than the previous one—9.173 syll/sec—but it doesn't sound quite as quick, perhaps because the previous utterance was longer and involved more percussive consonants.²⁴ The question to Chekhov is characterized by a lot of nearly-swallowed vowels.

²⁴ This is a good reminder that it isn't only short vowels and pauses that make speech seem quick. A flurry of plosives may work like clicks of a metronome.

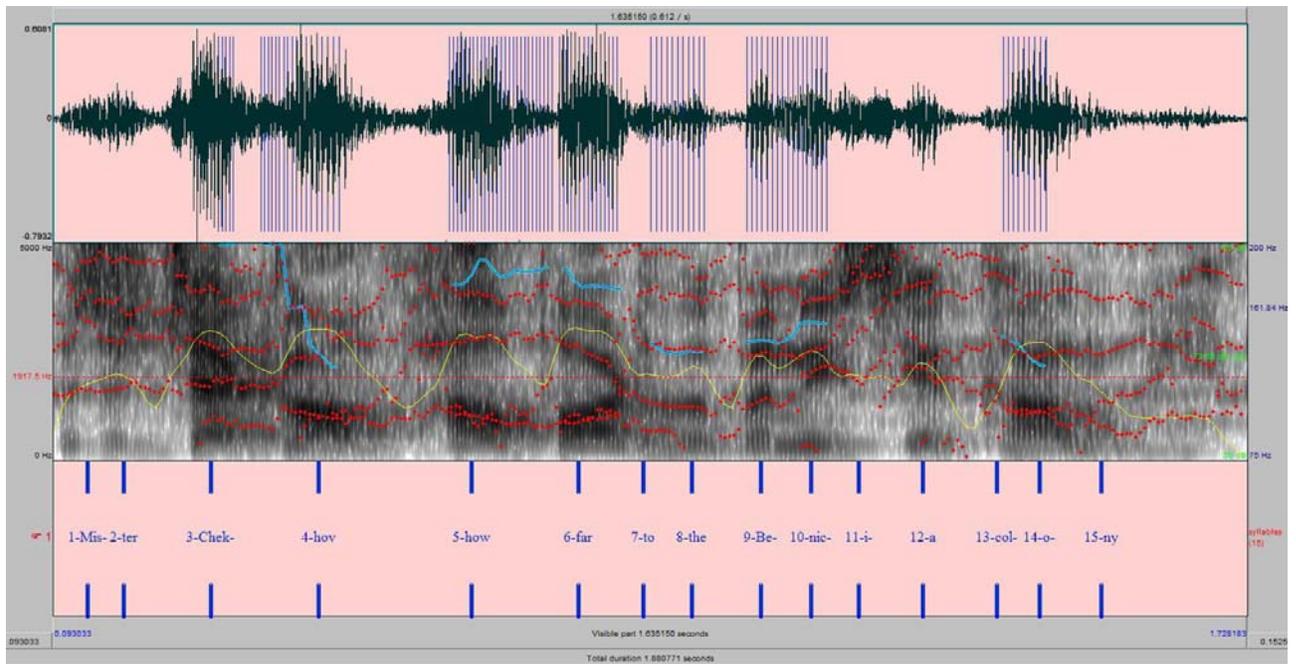


Figure 17: Lester-as-Kirk, “Mr. Chekhov, how far to the Benecia colony?” [http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman bridge 1b.wav](http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk%20as%20woman%20bridge%201b.wav)

In these utterances, Shatner is using a high pitch to convey that there’s a woman in Kirk’s body, but the tempo is also a crucial part of showing that it’s a hysterical woman who’s possessed him. Earlier I stated that men were seen to be faster talkers than women. This isn’t actually true (see, for example, Lalljee and Cook 1973) and I suspect the stereotype itself disappears in many situations.

John Wayne, a paragon of masculinity, had the following advice on acting (and perhaps more generally, *acting as a man*): “Talk low, talk slow, and don’t talk too much.” Shatner’s portrayal of Lester, by contrast, has him talking high, talking fast, and talking a lot. Men may be associated with fast talk, but mile-a-minute talk probably has different associations. We’ll need to add this to our map for the indexical field of tempo—there are various degrees of “fast talk”. Methodologically, we can compare absolute measurements but it’s probably even more useful to compare tempos within interactions.

The two helmsmen, Chekhov and Sulu have been contemplating mutiny right before Lester gets to the bridge, and here Chekhov’s response is faster than normal (5.008 syll/sec), but still doesn’t match Lester’s level. His relative slowness could be a sign of conversational defiance (“I won’t match your sense of urgency”), though my gut is that other acoustic features are conveying defiance more than the tempo.

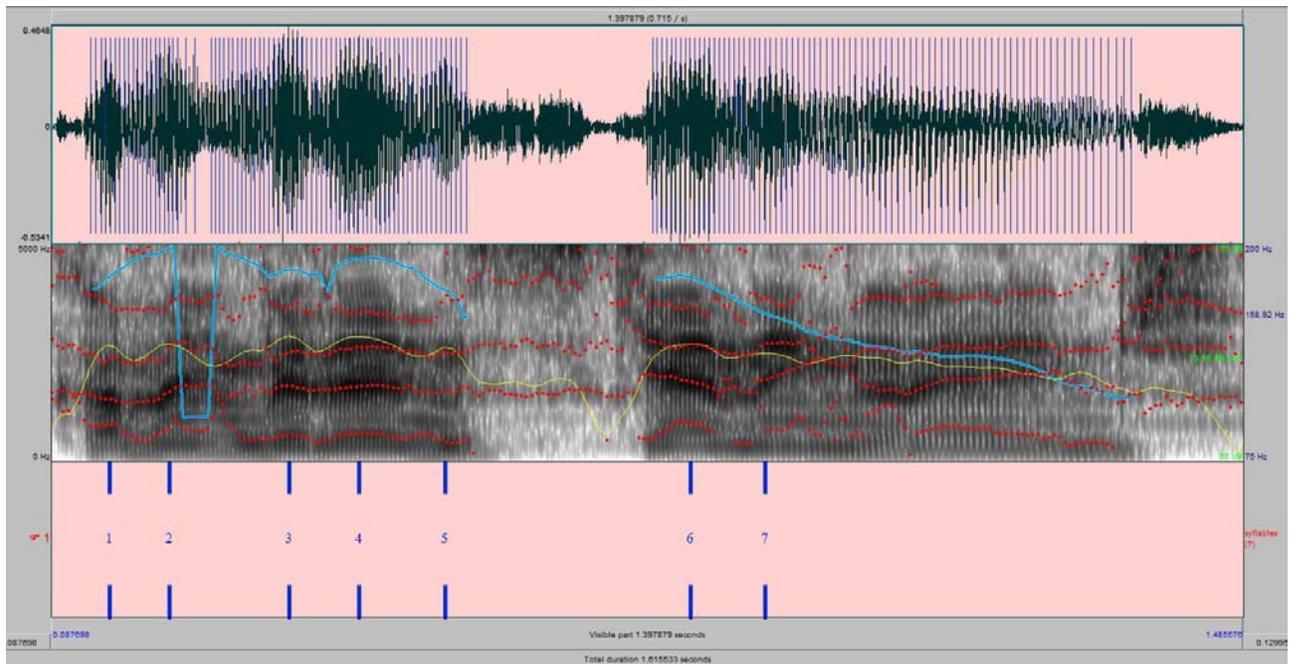


Figure 18: Chekhov replies, “Coming within scanning range.” 5.008 syll/sec. [http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman bridge 1c.wav](http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk_as_woman_bridge_1c.wav)

Shatner begins Lester’s next lines with panting,²⁵ which he does several times for Lester. The real Captain Kirk often inserts pauses but never pants. This is another signal of an aroused emotional state—Lester is out-of-breath and frantic. There is nothing dignified in her bearing and she is not worthy of the crew’s respect.

“Plot coordinates for orbit”, Lester says (4.443 syll/sec). Then a quick pause as she switches to address Sulu—“Mr. Sulu” (6.973 syll/sec)—a 0.335 second pause follows. Lester redoubles and accelerates to 7.224 syll/sec (“Lock in to coordinates as soon as orbit is accomplished”). She concludes by letting them know that “Interment will take place on Benecia” (4.883 syll/sec). There is a strong stress on “on” that slows that utterance down.

Though Shatner isn’t sounding bursty, he still is using several different tempos to make his speech stand out. In fact, it is probably because Shatner has different tempo levels that the extra-fast utterances sound so panicked. This recalls the finding in psychology and music that fear is expressed not only by fast tempo, but by irregular tempo. Emotional erraticism is conveyed by tempo irregularity.

²⁵ Not shown. [http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman bridge 1d.wav](http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk_as_woman_bridge_1d.wav)

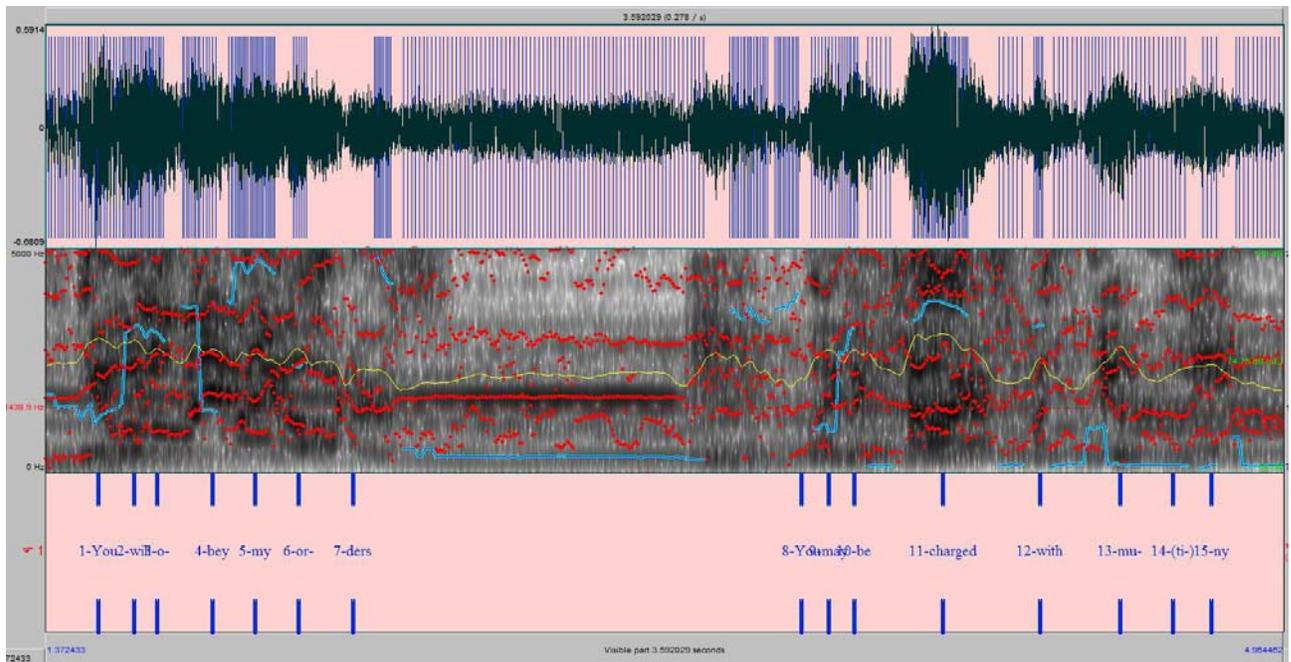


Figure 19: Lester-as-Kirk: “You will obey my orders...You may be charged with mutiny!” [http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman bridge 2.wav](http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk%20as%20woman%20bridge%202.wav)

Chekhov and Sulu put their hands in their laps. Lester is furious and she accelerates to 7.756 syll/sec (“You will obey my orders!”). After a 1.152 second pause, Lester levels the charge of mutiny, but this is said with more quiver in Shatner’s voice and the tempo slows to 5.747 syll/sec. There’s another uncertain pause—this time of 2.977 seconds—followed by a reiteration: “You will obey my orders, or...” This is delivered at 6.679 syll/sec, though the rush is dramatically interrupted as Kirk almost reclaims his body, thanks to the mental manipulations of Spock in the brig.

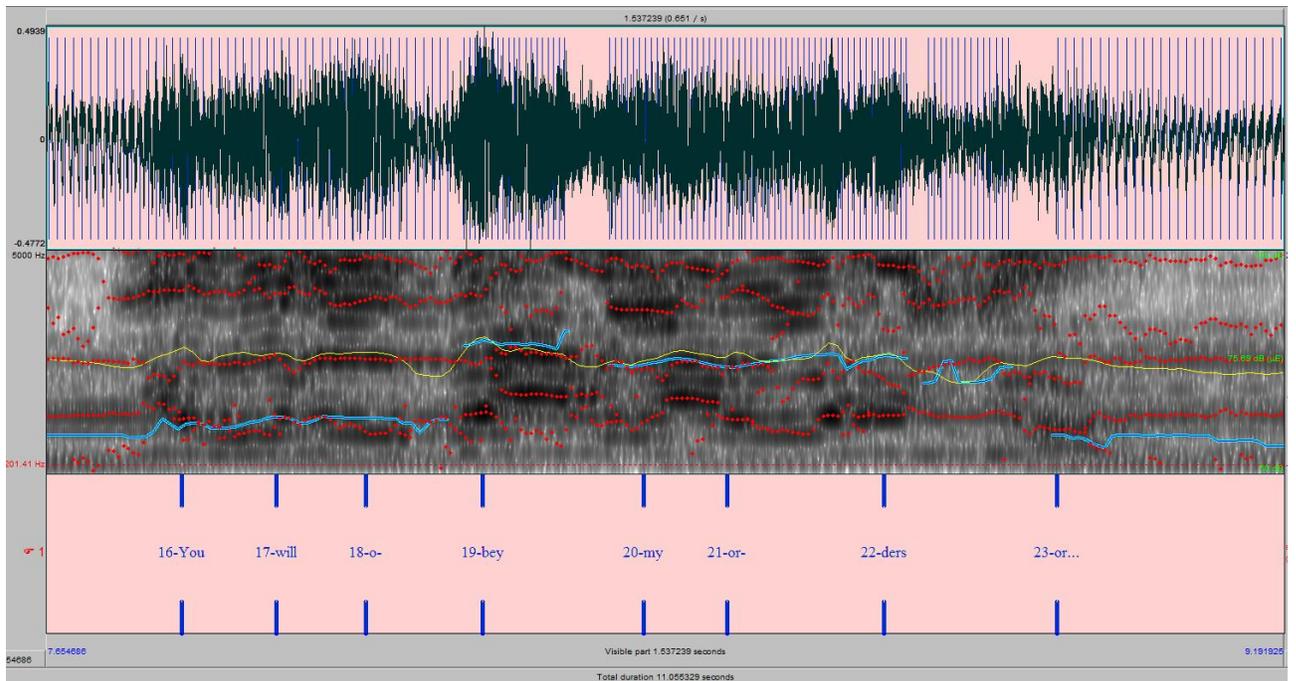


Figure 20: Lester-as-Kirk, “You will obey my orders, or...” <http://www.stanford.edu/~tvlers/notes/socioling/Sounds/Kirk as woman bridge 2.wav>

Lester manages to keep control of Kirk’s body and goes down to the brig to separate the conspirators. As Lester delivers the news, she sounds like she is performing “in control” for her audience of enemies and collaborators. There is still a quiver in her voice (and the pitch is still high), but the rate has slowed to 4.707 syll/sec.

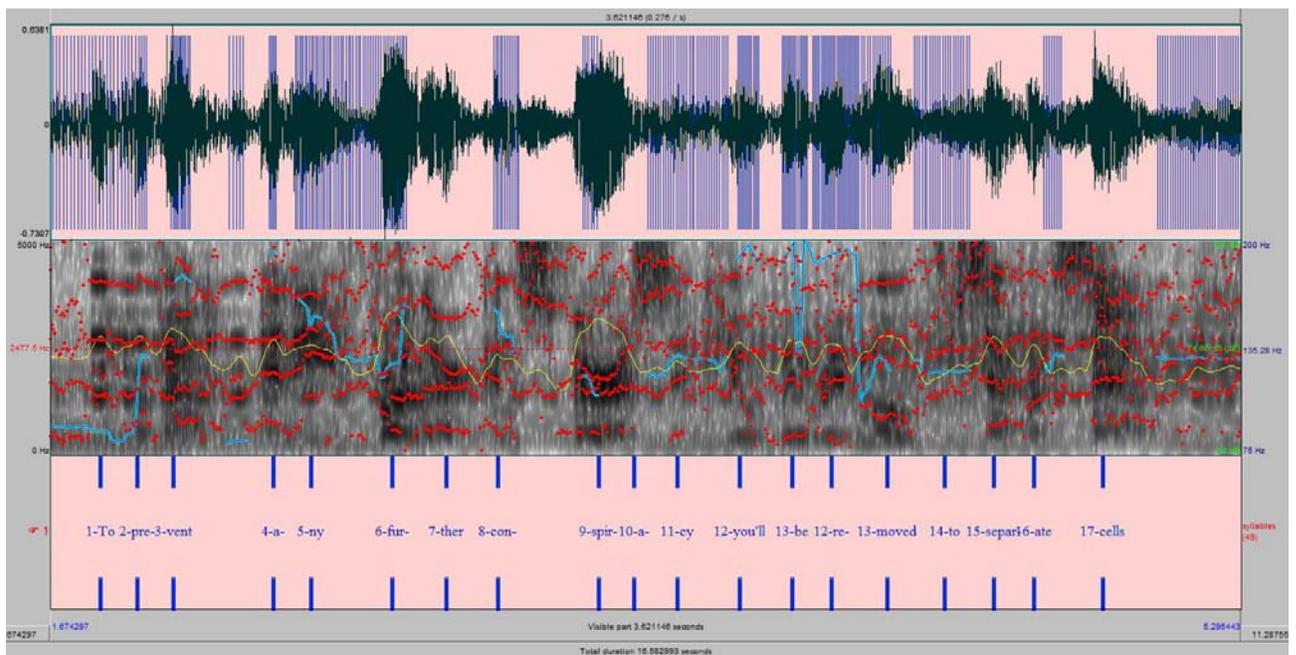


Figure 21: Lester-as-Kirk, “To prevent any further conspiracy, you’ll be removed to separate cells.” 4.707 syll/sec. <http://www.stanford.edu/~tvlers/notes/socioling/Sounds/Kirk as woman brig 3.wav>

In Lester’s next sentence we see a hint of Shatner burstiness. Lester is threatening Spock, McCoy, Kirk, and Scotty with sedatives, but there is a big dramatic pause in the midst of an “if” clause. The 0.504 sec pause begs the question that the rest of the sentence will deliver. It’s as if Shatner /Lester want their audiences to ask, “If there’s any resistance...well, what!?”).²⁶

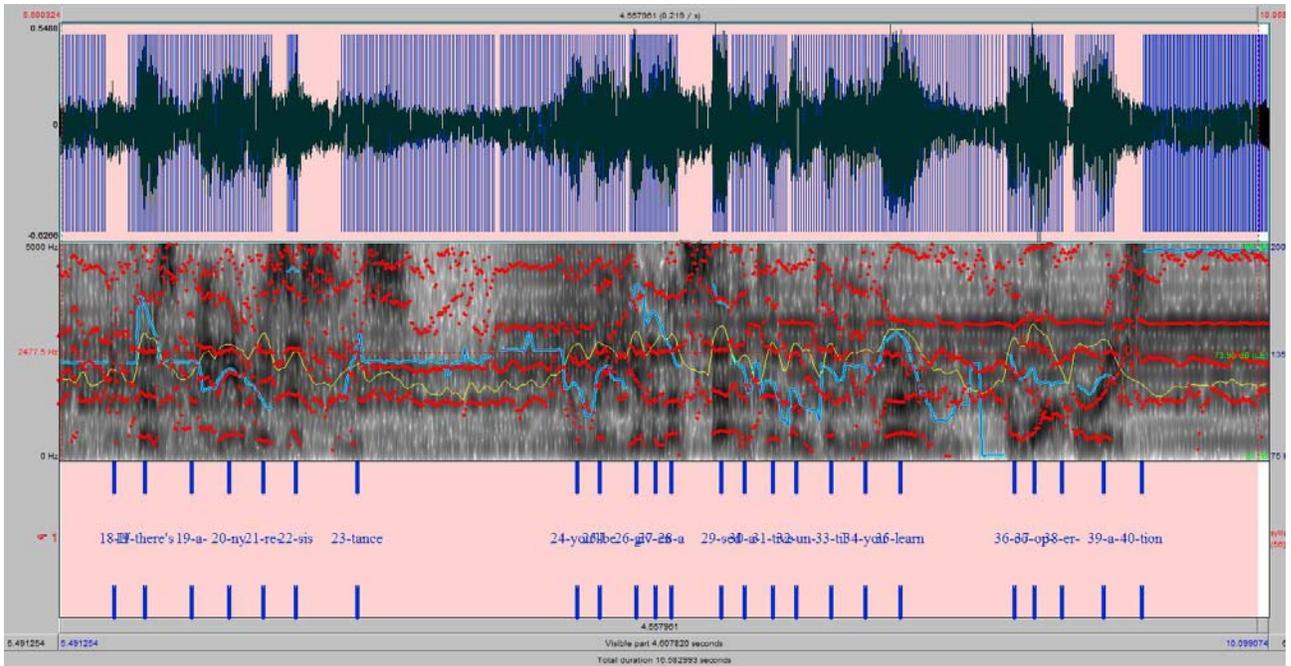


Figure 22: Lester-as-Kirk, “If there’s any resistance you’ll be given a sedative until you learn cooperation.” <http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman brig 3.wav>

As the prisoners leave their cells, there’s a struggle and Lester and Kirk touch. This reverses the body swap and now we hear Sandra Smith’s version of a hysterical Janice Lester.

²⁶ It might look like there’s a pause between “learn” and “cooperation”, but the /n/ leads right into the /k/.

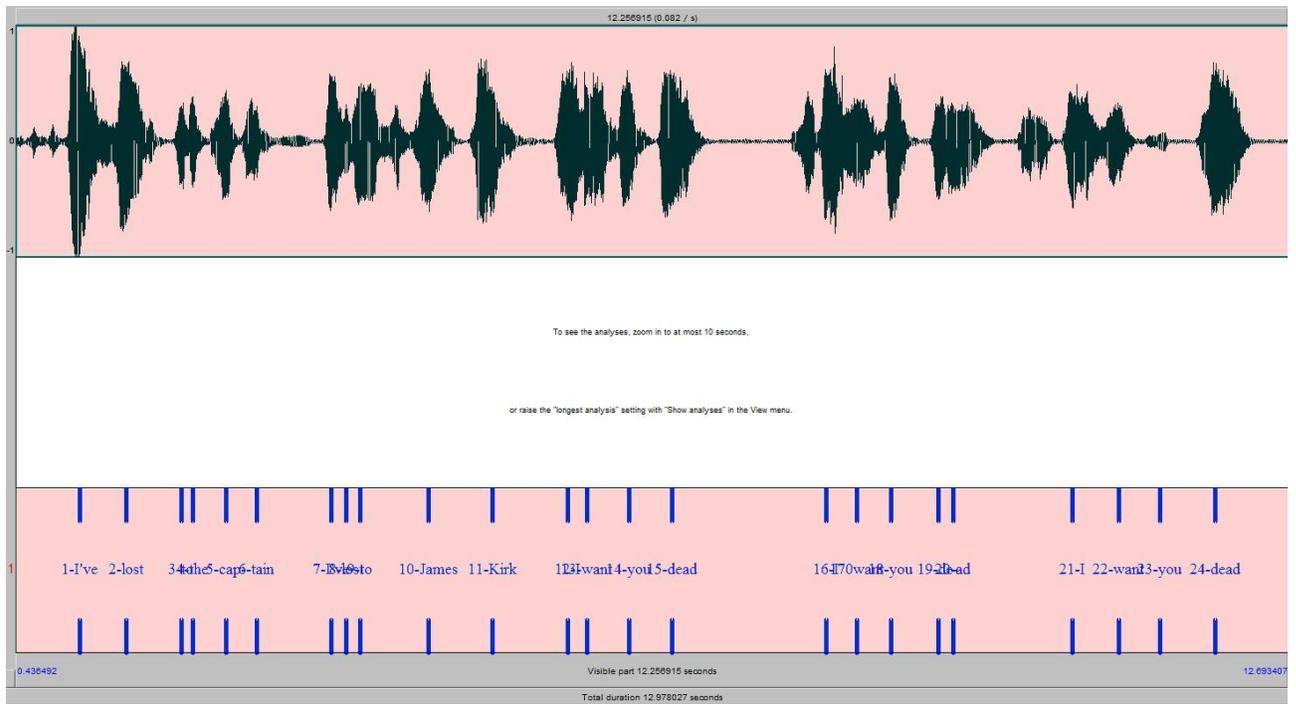


Figure 23: Lester-as-Lester, “I’ve lost to the captain. I’ve lost to James Kirk. I want you dead. I want you de-ad. I want you dead.” [http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk as woman Janice screaming 4.wav](http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk_as_woman_Janice_screaming_4.wav)

Her speech begins with two quick sobs. Overall, the speech rate is a slow 1.958 syll/sec, but part of that is due to the pauses. The first section is 2.375 syll/sec (“I’ve lost to the captain, I’ve lost to James Kirk”), followed by 2.444 syll/sec (“I want you dead”), then 2.500 syll/sec (“I want you de-ad”, where “dead” is turned into a bisyllabic word). There’s another sob and then 2.061 syll/sec as she moves into a softer, resigned, weeping “I want you dead.” The relative regularity of the tempo of these sections and their slowness foreshadows resignation. The arousal levels have plummeted.

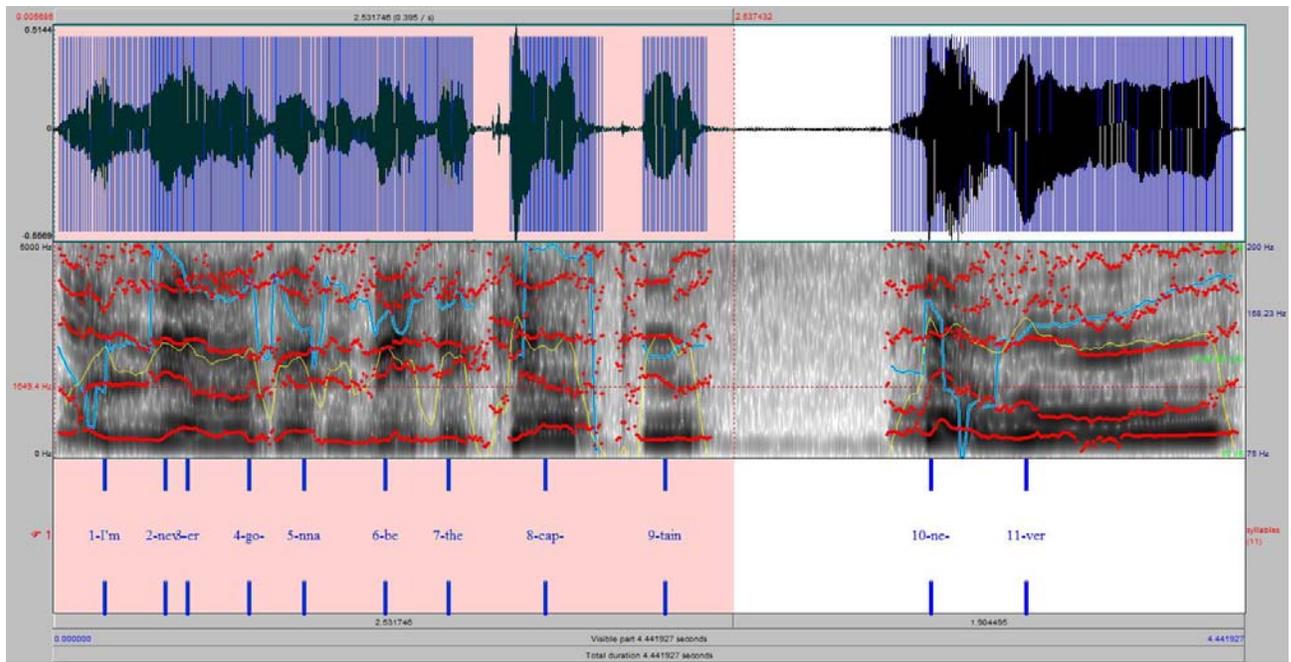


Figure 24: Lester-as-Lester, “I’m never gonna be the captain. Neverrrr...” http://www.stanford.edu/~tylers/notes/socioling/Sounds/Kirk_as_woman_Janice_screaming_5b.wav

Before collapsing into completely inarticulate moans, Lester descends into something that sounds like little girl self-pity. “I’m never gonna be the captain”, she laments (3.555 syll/sec). Then she repeats, “Never”, drawing the last syllable out for quite some time (the rate for this word would be 1.560 syll/sec).

The climax of the episode is really in the interactions in the courtroom and on the bridge. By the time we reach the brig, we’re pretty much in the episode’s denouement. We see that reflected in Lester’s tempo, both in Kirk’s body and her own. This is another way in which speech tempo recapitulates dramatic structure, building, hitting heights, slowing down.

Conclusion and next steps

In this paper, I have built up indexical fields for tempo and discussed the ideologies they demonstrate and help keep alive. I have done this by combing through psychology/musicology literature, searching various corpora, and running a pilot survey of 50 people across the country for their answers to who talks fast and who talks slow.

I described some of the opportunities that acted scenes present us for studying style (and tempo in particular). This involved a discussion of performance and the actor’s role. It also brought me towards a close analysis of several scenes in *Star Trek* that feature William Shatner. Shatner varies his tempo a great deal and I hope reading of the scenes will encourage a closer look at tempo variation and the contextual/interactional aspects of style.

My approach of using corpora, surveys, prior research, and intuitions is not fool proof—I may miss patterns that are really there even though people don’t comment upon them or study them. The fact is that we’re still in the infancy of stylistic cosmography.

As constellations are proposed for different variables, we will want to see how the meanings for one variable map to the meanings of another. This will expand beyond explicitly mentioned meanings. In visualizing the indexical field of /t/ release, for example, Eckert (2008) includes elements like “emphatic” and “angry”, which we’ll see in our “fast talk” indexical field. Eckert also puts in “Nerd Girl” and “Gay Diva”. I haven’t put these in the “fast talk” indexical field, but they may be worth exploring. The “fast talk” field doesn’t have “educated”, either, but it does have “intelligent” and “knowledgeable”—“educated” has an ideological connection to British and school-teacher for /t/ release, while the intelligence and knowledgeability of fast talkers is more likely to be linked to persuasion. Meanwhile, some meanings for /t/-release connect to “slow talk”—“careful” and “polite”, for example. How are different fields related to each other? As we fill out more indexical fields, we’ll want to keep an eye out for such similarities and contradictions.

Future work will, of course, refine the indexical fields I have put forward here. Some specific directions:

- Extend the pilot survey
- Further explore tempo in *Star Trek* and other William Shatner vehicles, building beyond close analysis to overall aggregation²⁷
- Add scenes from other people known to have bursty speech—Christopher Walken, for example, and Winston Churchill
- Verify that my proposed measure of burstiness (*variance/mean*) is the best one to use
- Consider production and perception experiments to augment the analysis of scenes and the charting of indexical fields

²⁷ I’ve done this to some extent, but this paper was already lengthy so it didn’t seem appropriate to add any more scenes here or even to add a section that aggregates everything.

Works cited

- Agar, M. 1994. *Language shock: Understanding the culture of conversation*. New York: William Morrow and Company. <http://bartleby.com/66/54/3154.html>
- Apple, W., Streeter, L.A., & Krauss, R.M. 1979. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37, 715–727.
- Averill, J. 1980. A constructivist view of emotion, in R. Plutchik and H. Kellerman (eds.). *Theories of Emotion*, New York, Academic Press, 306-329.
- Bachorowski, J.A., 1999. Vocal expression and perception of emotion. *Current Direct. Psychol. Sci.*, 8 (2), 53-57.
- Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.*, 70 (3), 614-636.
- Butler, J. 1988. *Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory*. *Theatre Journal*, 49 (1), 519-531.
- Brown, B., H. Giles, and J. Thackerar, 1985. Speaker evaluations as a function of speech rate, accent and context. *Language and Communication*, 5, 207-220.
- Brown, B., W. Strong, and A. Rencher. 1973. Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *Journal of the Acoustical Society of America*, 54, 29-35.
- Brown, B., W. Strong, and A. Rencher. 1974. Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings from speech. *Journal of the Acoustical Society of America*, 55, 313-318.
- Caine, M. 1992. Interview in *The Times* of London. 26 Aug 1992. <http://www.bartleby.com/66/20/10020.html>
- Clark, H. 2002. Speaking in time. *Speech Communication*, 36, 5-13.
- Collins, S. 1989. Subjective and autonomic responses to Western classical music. Unpublished doctoral dissertation, University of Manchester, UK.
- Cook, M. 1969. Anxiety, speech disturbances and speech rate. *Brit. J. Soc. Clin. Psycho*, 8, 13.
- Cowie, R and R. Cornelius. 2003. Describing the emotional states that are expressed in speech, *Speech Communication*, 40 (1-2), 5-32.
- van Dommelen, W. and B. Moxness. 1995. Acoustic Parameters in Speaker Height and Weight Identification: Sex-Specific Behaviour, *Language and Speech*, 38:3, 267.
- Eckert, P. 2008. Variation and the indexical field. *Journal of sociolinguistics* 12 (3), 453-476.
- Ekman, P., 1992. An argument for basic emotions. *Cogn. Emot.* 6 (3/4), 169-200.

- Fairbanks, G. and L. Hoaglin. 1941. An experimental study of the durational characteristics of the voice during the expression of emotion. *Speech Monographs*, 8, 85.
- Feldstein, S. F. Dohm, C. Crown. 1993. Gender as a mediator in the perception of speech rate. *Bulletin of the Psychonomic Society*, 31, 521 - 524.
- Goldman-Eisler, F. 1968. *Psycholinguistics. Experiments in Spontaneous Speech*, London, Academic Press.
- Hart, R. 1971. Personality information contained in the vocal qualities and content aspects of speech. Unpublished master's thesis, Brigham Young University.
- Heuft, B., T. Portele, T., and M. Rauth. 1996. Emotions in time domain synthesis. In: *Proc. ICSLP 96*, Vol. 3, Philadelphia, USA, 1974-1977.
- Hevner, K. 1937. The affective value of pitch and tempo in music. *American Journal of Psychology*, 49, 621-630.
- Iida, A., N. Campbell, S. Iga, F. Higuchi, and M. Yasumura. 1998. Acoustic nature and perceptual testing of corpora of emotional speech. In: *Proc. ICSLP 98*, Sydney, Vol. 4, 1559-1562.
- Irvine, J. 2001. "Style" as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. Rickford (eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 21-43.
- Johnstone, T., Scherer, K.R., 2000. Vocal communication of emotion. In M. Lewis, M. and J. Haviland (eds.), *Handbook of emotion*, second ed.. Guilford, New York, 220-235.
- Kanfer, F. H. 1960. Verbal rate, eyeblink and content in structured psychiatric interviews, *J. Abnorm. Soc. Psychol.*, 61, 341.
- Kimble, C., J. Yoshikawa, H. Zehr. 1981. Vocal and verbal assertiveness in same-sex and mixed-sex groups. *Journal of Personality and Social Psychology*, 40, 1047-1054.
- Ladd, D., K. Silverman, F. Tolkmitt, G. Bergmann, K. Scherer. 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *J. Acoust. Soc. Amer.* 78, 435-444.
- Lalljee, M. G., & Cook, M. 1973. Uncertainty in first encounters. *Journal of Personality and Social Psychology*, 26, 137-141.
- Lalljee, M. G. and M. Williams. 1967. Speech rate and speech disturbance in the perception of anxiety. Unpublished paper, University of Oxford.
- Mannheim, B. 2000. Iconicity. *Journal of Linguistic Anthropology* 9:103-106. Reprinted (2001) In A. Duranti (ed.), *Key terms in language and culture*, Oxford: Blackwell. 102-105
- Miller, N., G. Maruyama, R. Beaber, and K. Valone. 1976. Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 34, 615-624.

- Nietzsche, F. 1886. *Beyond Good and Evil, Fourth Part: Maxims and Interludes*, section 85. In Friedrich Nietzsche, G. Colli and M. Montinari (eds.), Berlin, de Gruyter, 1980.
- Paivio, A. 1963. Audience influence, social isolation and speech, *J. Abnorm. Soc. Psychol.*, 67, 247.
- Ray, G. B. 1986. Vocally cued personality prototypes: An implicit personality theory approach. *Communication Monographs*, 53, 266-276.
- Ray, G. and C. Zahn. 1999. Language attitudes and speech behavior: New Zealand English and Standard American English. *Journal of Language and Social Psychology* 18.310–319.
- Ray, G. B., E. Ray, and C. Zahn. 1991. Speech behavior and social evaluation: An examination of medical messages. *Communication Quarterly*, 39, 119-129.
- Scherer, K. R. 1981. Speech and emotional states. In J. Darby (ed.), *Speech evaluation in psychiatry*. New York: Grune & Stratton, 189-220.
- Scherer, K.R., 1986. Vocal affect expression: A review and a model for future research. *Psychol. Bull.* 9 (2), 143-165.
- Scherer, K. 2003. Vocal communication of emotion: a review of research paradigms, *Speech Comm.* 40 227–256.
- Scherer, K.R., Oshinsky, J.S., 1977. Cue utilization in emotion attribution from auditory stimuli. *Motiv. Emot.* 1, 331-346.
- Silverstein, M. 1976. Shifters, linguistic categories, and cultural description. In *Meaning in anthropology*, K. Basso and H. Selby (eds.). Albuquerque: University of New Mexico Press, 11-55.
- Smith, B, B. Brown, W. Strong, and A. Rencher. 1975. Effects of speech rate on personality perception. *Language and Speech*, 18, 145-152.
- Steer, A. 1974. Sex differences, extraversion and neuroticism in relation to speech rate during the expression of emotion , *Language and Speech*, 80.
- Street, R. 1985. Participant-observer differences in speech evaluation. *Journal of Language and Social Psychology* 4.125–130.
- Street, R. L., and R. Brady. 1982. Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs*, 49, 290-308.
- Streeter, L. R. Krauss, V. Geller, C. Olson, and W. Apple. 1977. Pitch changes during attempted deception. *J Pers Soc Psychol.*, 35 (5), 345-50.
- Ververidis, D. and C. Kotropoulos. 2006. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48 (9), 1162–1181.

Appendix A: Slow talk indexical field

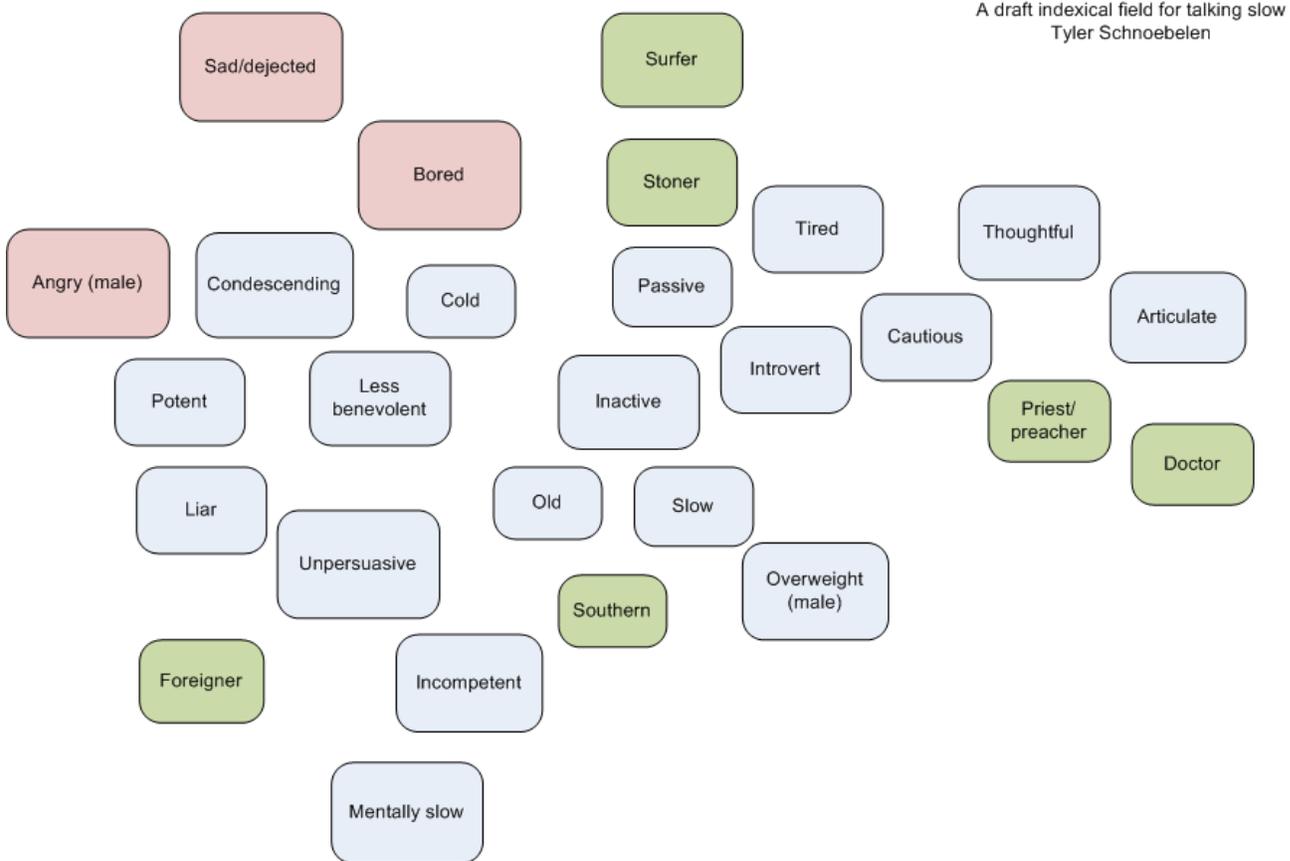


Figure 25: The indexical field for slow talking, built from literature review, corpora searches, and survey results.

Appendix B: Indexical fields as lists (with references and some notes)

Fast speech

- Pakeha New Zealander²⁸
- Rage/hot anger²⁹
- Fear/panic/terror³⁰
- Joy/elation³¹
- Anxiety³²
- Extraversion
- Maleness³³
- Honesty³⁴
- Fluency³⁵
- Persuasiveness³⁶
- Less potency³⁷
- Intelligence³⁸
- Less benevolence³⁹
- Objectiveness⁴⁰
- Competency⁴¹
- Emphaticness⁴²

²⁸ Ray and Zahn (1999) is one of the few sociolinguistic studies I found investigating the social meaning of tempo. ‘Pakeha’ indicates New Zealanders of European descent. Ray and Zahn interpret it as “upgraded status” compared to Maori/Polynesian status (slower speech, “downgraded status”). I didn’t know where to place this in my indexical field, so I left it off.

²⁹ Johnstone and Scherer (2000), reported in Scherer (2003).

³⁰ Johnstone and Scherer (2000), reported in Scherer (2003). See also Scherer (1986) and Banse and Scherer (1996)

³¹ Not completely clear-cut. Johnstone and Scherer (2000), reported in Scherer (2003); see also Scherer (1986) and Banse and Scherer (1996). “While joy is almost perfectly recognized in the face, listeners seem to have trouble identifying this emotion unequivocally in the voice” (Scherer 2003: 235).

³² Some people speed up when they’re anxious, others slow down (Cook 1969). See also Kanfer (1960) and Paivio (1963).

³³ Feldstein, Dohm, and Crown (1993). Women tend to judge everyone as speaking faster than they are. But both women and men believe that male speakers talked faster than the females, even though the speech rates were comparable. Laljee and Cook (1973) also found no significant sex difference in speech rate.

³⁴ Apple et al (1979).

³⁵ Apple et al (1979).

³⁶ Apple et al (1979), Miller et al (1976). Apple et al find that talking quickly on some subjects does lower persuasiveness compared to a normal rate. They compared people talking about affirmative action (faster=more persuasive) versus how to budget (faster=less persuasive).

³⁷ Apple et al (1979).

³⁸ Miller et al (1976).

³⁹ Brown et al (1973), Brown et al (1974). Really benevolence has a U-shaped curve where the normal rate voices get the highest scores. The data here is somewhat complex and this item may not belong on the chart.

⁴⁰ Miller et al (1976).

⁴¹ Smith et al (1975), Brown et al (1973), Brown et al (1974). See also Ray (1986), Ray, Ray and Zahn (1991), Street and Brady (1982). Ray and Zahn (1999) propose that competence is comparable to “power”. For Ray and Zahn, competence is linked to specific questions about speaker competence as well as questions about speaker carefulness and education. For Smith et al (1975), competence is made up from questions about activeness, ambitiousness, intelligence, looks, confidence.

⁴² Apple et al (1979).

- Knowledgeability⁴³
- Credibility⁴⁴
- Social attractiveness (friendly, warm, caring, good-natured, emotional, polite)⁴⁵
- Dynamism (talkative, dominant, headstrong, active, extroverted)⁴⁶

Slow speech

- Sadness/dejection⁴⁷
- Boredom⁴⁸
- Subdued/cool anger (male)⁴⁹
- Overweight (male)⁵⁰
- Introversion
- Dishonesty⁵¹
- Lack of fluency⁵²
- Lack of persuasiveness⁵³
- Passivity (inactive, slower, colder, passive, weaker)⁵⁴
- Potency⁵⁵
- Less competency⁵⁶
- Less benevolence⁵⁷
- Less emphaticness⁵⁸

⁴³ Miller et al (1976).

⁴⁴ Miller et al (1976).

⁴⁵ For moderately fast rates: too fast or too slow meant lower social attractiveness. Ray and Zahn (1999) give references to Ray (1986) and Ray et al (1991). R&Z consider “social attractiveness” to be similar to “solidarity”.

⁴⁶ Ray and Zahn (1999).

⁴⁷ Johnstone and Scherer (2000), reported in Scherer (2003). See also Scherer (1986) and Banse and Scherer (1996).

⁴⁸ Johnstone and Scherer (2000), reported in Scherer (2003).

⁴⁹ Steer (1974) finds male anger to be slower than female anger. See also Lalljee and Williams (1967) and Fairbanks and Hoaglin (1941).

⁵⁰ van Dommelen and Moxness (1995). References for people trying to detect sex, age, race, too. See also their definition of articulation, pg 5. Males were able to detect, females weren't. Female weight couldn't be detected in any consistent pattern. Active/inactive, inactive=passive=fat. Also, what does it mean to be overweight and think of oneself as such.

⁵¹ Apple et al (1979) for the perception; Streeter et al (1977) found that when they gave subjects motivations to lie, they spoke more slowly, when they just asked them to lie, they spoke more quickly. The effects for the “with instructions/motivation” were weak. Perhaps lying is assumed to take an extra bit of cognitive energy, slowing down the rate?

⁵² Apple et al (1979).

⁵³ Apple et al (1979), Miller et al (1976).

⁵⁴ Apple et al (1979).

⁵⁵ Apple et al (1979).

⁵⁶ Smith et al (1975), Brown et al (1973), Brown et al (1974).

⁵⁷ Smith et al (1975); benevolence scores decrease whenever you go above or below a normal pace. Hart (1971) cautions that while competence judgments use vocal qualities and verbal content, benevolence judgments seem to be based mainly verbal content.

⁵⁸ Apple et al (1979).

Appendix C: Survey results

I ran a quick pilot survey on Amazon’s Mechanical Turk to see what sort of people “talk fast” or “talk slowly”. The majority of my participants were women in their 20s or 30s with at least a BA. Few of them were from the West Coast—but otherwise, the participants covered a nice geographical area. For reasons of length, I put off the interesting discussion of how age/gender/region mapped to answers. For example, did women in their 20s say that “teenage girls” talk fast? Did Southerners say that they themselves “talk slow”?

Since these were open-ended questions, I got a variety of associations. For ease of reading, I created labels to group like responses together.

Who talks fast?	Count
Anxious	12
Salesperson	10
Young	8
ADD	5
Kid	5
Liar	5
Extrovert	5
Auctioneer	5
Intelligent	5
Politician	4
Excited	4
Advertiser	4
Teacher	4
Happy	3
Rushed	3
Spanish speaker	3
New Yorker	3
Con-man	2
Rote speakers	2
Stressed	2
Urban	2
Authoritative	2
Northeasterner	2
Busy	2
People in finance	2
Asian	2
Female	2
News reporter	2
Young (female)	2
Italian	2
Hispanic	2
Waitress	1
African-American rapper	1
Energetic	1
Acting smart	1

Female (ditsy)	1
Impatient	1
High-strung	1
Indian	1
Radio host	1
Inexperienced	1
Cocaine user	1
Communicator	1
Foreign speaker	1
Angry	1
Grew up with fast talking family	1
Japanese girls	1
Political pundit	1
Customer service	1
Public speaker	1
Lawyer	1
Auto-mac operator	1
Debater	1
Fast food worker	1
DJ	1
Sportsman	1
Doctor	1
Studious	1
East coasters	1
French	1
Northerner	1
Want to hold the floor	1
Passionate	1
Comedian	1
Egotistical	1
Police officer	1
Grand Total	151

Figure 26: Who talks fast? Survey results from 50 people.

Who talks slowly?	Count
Southerner	16
Old	13
Teacher	10
Mentally slow	9
Tired	5
Preacher/priest	5
Doctor	5
Foreign speaker	4
Farmer	4
Speaking for	3

comprehension	
Kid	3
Nurse	2
New Age	2
Precise	2
Relaxed	2
Newscaster	2
Counselor	2
Intelligent	2
Uncertain	2
Boring	2
Sick	2
Mature	2
Speaking to children	2
Trying to control emotion	2
Bored	2
Romantic	1
Fill time	1
Speaking to foreigner	1
Computer scientist	1
Redneck	1
Irish	1
Grandparent	1
Confident	1
Politician	1
Librarian	1
Public speaker	1
Male	1
Caregiver	1
Boss	1
Sad	1
Mechanic	1
Hillbilly	1
Customer service	1
Bank teller	1
Deaf	1
Female	1
Depressed	1
Professional	1
Calm	1
Radio host	1
Drunk	1
Refined	1
Non-native	1
Reserved	1
Canadian	1

Russian	1
African-American	1
Shy	1
Old (men)	1
Athlete	1
Patient	1
Important	1
Physicist	1
Stoner	1
Police officer	1
Insecure	1
Polish	1
Wise	1
Western	1
Methodical	1
Administrator	1
Midwestern	1
Grand Total	152

Figure 27: Who talks slowly? Survey results from 50 people.

Demographics

Here, for reference, is the demographic information.

Grew up	Count
East	12
Midwest	10
International/other	7
South	15
West	6
Grand Total	50

Age	Count
19-29	20
30-39	15
40-49	9
50+	6
Grand Total	50

Gender	Count
Female	38
Male	12
Grand Total	50

Education	Count
-----------	-------

Schnoebelen

High school graduate	3
Some college, no degree	7
Associates degree	4
Bachelors degree	26
Graduate degree (Masters, Doctorate, etc.)	10
Grand Total	50

Appendix D: Music and emotion

What passion cannot Music raise and quell? – Dryden (1687)

Findings for speech rate look a lot like what we see in results for tempo in music.

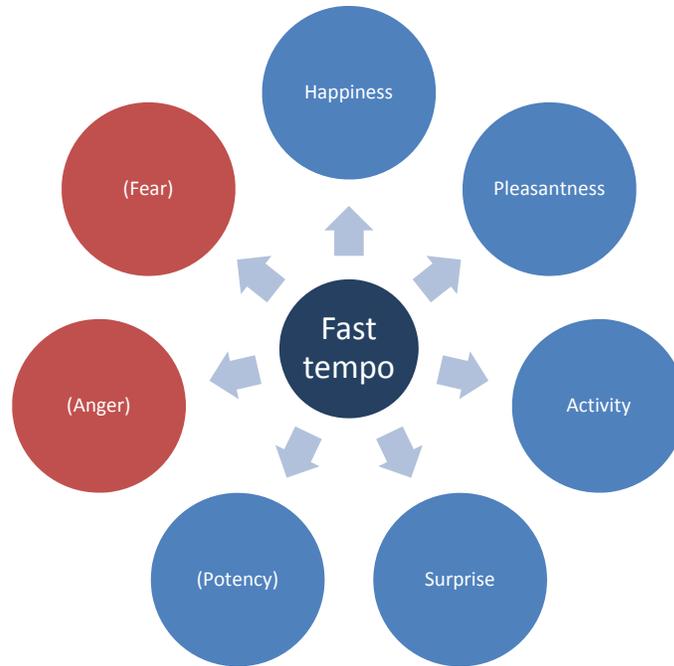


Figure 28: Emotional states corresponding to fast tempo (Scherer and Oshinsky 1977: 340). Those in parentheses had other characteristics ranked higher than tempo. For example, "anger" is expressed mainly through many harmonics and then by fast tempo.

Scherer (1981: 206) adds confidence and indifference; Collins (1989: 45) would have us add excitement but probably wouldn't put up happiness. Fear is especially associated with a highly irregular tempo (Collins 1989: 45).

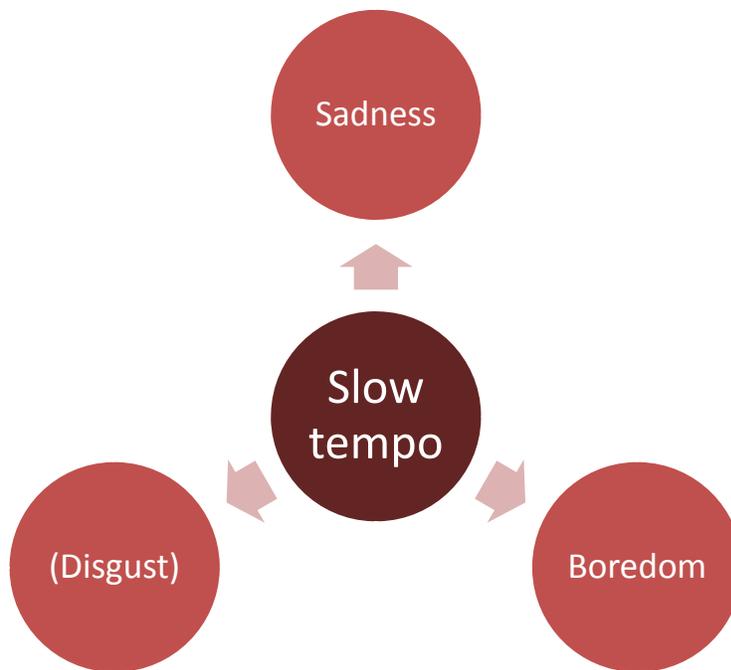


Figure 29: Emotional states corresponding to slow tempo (Scherer and Oshinsky 1977: 340). Slow tempo is a rather minor part of what makes music signal “disgust” (it’s more about (i) many harmonics, (ii) small pitch variation, (iii) round envelope).

For slow tempo, Scherer (1981: 206) adds contempt. We may also want to add tenderness, solemnity and/or relaxation (Collins 1989: 45 for the last one). In the older literature, Hevner (1937: 624) reports that slow tempos (63, 72, or 80 bpm) are “dignified...and calm-serene”, while fast tempos (102, 104, 112, 152 bpm) are “happy-gay and exciting restless”.