UC Merced
UC Merced Undergraduate Research Journal

Title
An Investigation of Hand Gestures and its Application to Rhythm and Semantics

Permalink
https://escholarship.org/uc/item/8kn1s6qx

Journal
UC Merced Undergraduate Research Journal, 8(1)

Author
Ramirez, Adolfo

Publication Date
2015

DOI
10.5070/M481029504

Undergraduate

THERE IS NO

SUCH THING

AS BAD

RESEARCH

An Investigation of Hand Gestures and its Application to Rhythm and Semantics:

Actions That Come in Handy

By Adolfo G. Ramirez

Abstract

This study specifically focuses on the rhythmicity of hand gestures used as emphasis in speeches, in which either a public speaker is talking to a silent audience or a to a responding audience, and whether emphasis in these cases can be said to match our expectations of aligning with words that carry higher semantic weight. The data observed is the frequency of hand gestures aligning with either function or lexical words. This data gives a deeper look into the predictability of emphasis that expands to not only phonology but to motor functions such as hand gesturing. It is found that hand gestures align more frequently with lexical words than function words, which is to be expected, but surprisingly a decent number of function words are also being emphasized. Although function words carry very little semantic weight, their frequency of hand gesture lineups seem to reveal that they are valid targets for emphasis in a situation in which it is adjacent to other lexical words being emphasized. In the case of an isolated function word emphasis, these function words seem to carry crucial semantic information for the sentence although the word itself has a relatively low semantic weight. Such data gives more evidence for the integration of various motor functions such as hand gestures in speech production, to work as a facilitator of metaphysical information in which its ties to rhythm may be limited.

The use of hand gestures is a common practice for every public speaker, whether it'd be lecturing in front of a classroom or chatting with your friend over a cup of coffee. It is common knowledge that someone who moves their hands in a conversation is perceived to be more engaged in the topic. In debates and expositions of important topics, hand gestures are very prominent and can be used to make emphasis about certain points. Emphasis, in terms of rhythm, can be said to be a peak that stands out in a specific pattern. Accents in language, sound intensity in music, and brisk movements in dancing are just a few examples of emphasis in rhythm. Hand gestures have a peculiar way of lining up with our speech processes and raise the mystery of whether this use of emphasis can be linked as a piece of evidence for the existence of rhythm in speech.

The study of multimodal processes speech processes has been a study of interest that has increasingly become more accessible over time with new and better recording equipment and speech analysis tools. To be multimodal with communication has been a broad term in various disciplines, but it generally refers to the plurality of functions that seem to synchronize or work in reference to each other to produce some output, which in this case becomes speech and gestural emphasis in the English prosody. (Granstrom et al., 2013) It is possible for us to understand the tone and emphasis of certain words through only hearing someone speak. That extra layer of semantics may be lost when only looking at a transcript. Hand gestures add on to the semantic space by physically indicating emphasis on specific words. So, would it be easier to understand someone we can see and hear who also makes gestures in comparison to someone who simply speaks? Intuitively, it makes sense to attribute hand gestures as a facilitator for specific information, and therefore making it easier to understand someone. With this observation, multimodality in speech processes is described to work as an integrated system of information in which motor functions such as hand gestures work side by side with the semantic space of any given speech. In this investigation, such multimodality of prosody and motor functions such as hand gestures are found to have a relationship that affects one another.

The embodied and dynamic cognition movement in Cognitive Science look at functions as belonging to an integrated system that when put together have a unique emergent property that could not be seen when looking at the individual parts of a function. This idea was taken in a study by Glenberg and Gallese in which they use the Wolpert et al. MOSAIC model for motor movements and try to modify it to explain language as a piece of a bigger integrated system of communications. This model breaks up motor movements into two primary parts, a controller and a predictor. The controller is the primary driver for movement initiation and the predictor is juxtaposed to work in a parallel fashion when it comes to giving data to the controller for optimal movement. (Glenberg et al. 2012) Simply put, this model can be used to describe simple and complex motor tasks to great detail. With some simple modification for language, the model was used to describe and predict motor movements adjunct to speech production. The idea behind this comparison was to show that speech production can be made sense of as another adjunct layer in a motor function hierarchy. (Glenberg et al. 2012) This means that multimodal functions in speech processes can work in coordination to transform and process information.

From an evolutionary standpoint, scientists have compared communication and rhythm in monkeys to that of humans. Ghazanfar and Takahashi are scientists who focused on the Macaque monkey and the bi-modal use of facial gesture and lip smacking, which is interpreted to be evidence for a codependence in rhythm and language production. (Ghazanfar et al., 2012) In this light, humans are uniquely separated from the Macaque in communication because our speech

production is not so exclusively tied in with multi-sensory functions i.e., we can still talk over the phone to our friends and family while making gestures without seeing them and can still communicate through just movements alone and so on. For the Macaque, the bi-modal use of facial gesture and lip smacking may be a relationship crucial for the development of language, but when looking at a species that already developed language usage, it seems that the relationship's exclusiveness may die down. Despite that, it is further evidence for the way coordination between multiple motor functions may serve a purpose for an emergent property.

A study by Winter et al. studied a corpus of news archive videos of experienced public speakers and their use of gestures in relevance to how it lined up with the meaning in their speeches. They found that certain hand movements and gestures seemed to physically embody certain metaphors about quantity and numbers from their speech. For example, if a speaker was making a point about a certain statistic that seems to be growing, the speaker would line it up with a hand gesture highlighting the large size of it and then from left to right increasing the size of the gesture.(Winter et al, 2013) These behaviors seem to line up to the way we organize our thought and are used to emphasize the semantic space during a presentation. This is one instance it which it becomes known that speakers' physical gestures are directly coordinated with the semantic space of their presentation. In this study, the coordination of hand gestures and specific type of words is studied to further analyze how such coordination may be predicted and what that relationship may say about the semantics of the type of words involved in such coordination.

The investigation sampled two engaging YouTube videos on trending debate topics in which one was categorized as an interactive interview/debate and the other video was categorized to represent a unidirectional speech. Both of the topics were engaged in talk about the United States' economy and science education funding vs. conservative political agendas. The reason for keeping the topics similar was so that there would be minimal variability with possible confounds since the sample size is small. The first video is by the Detailed Report channel titled "Bill Nye Schools a Republican Congresswoman on Climate Change" in which scientist and public figure Bill Nye is in an interview on NBC debating with congresswoman

Marsha Blackburn. The video goes back and forth with the host of the interview David Gregory, Bill Nye, and Marsha Blackburn; due to the nature of the video, it is used to represent interactive speeches. On the other hand, the second video was by the channel Big Think titled "Neil deGrasse Tyson: Bringing Commercial Space Fantasies Back to Earth" in which scientist and public figure Neil deGrasse Tyson speaks by himself to the online audience. This video is meant to be a sample that represents unidirectional speeches and is compared to the interactive sample. The comparison was chosen so as to better understand whether there is a significant difference in these different kinds of speech presentations, and whether the data would point in different areas for each case. This type of comparison allows for a better generalization of the results, especially with a small sample size and could probably give more information of the patterns found in different types of speeches and hand gesturing. The videos chosen were under three minutes, with the Neil deGrasse Tyson video lasting two minutes and forty-one seconds and the Bill Nye video lasting two minutes and twenty-eight seconds. The comparison lies on the relative percentage of lexical versus function word to hand gesture line up rather than just frequency total, so the small difference in video length should not skew the data.

The key data extracted from the different videos includes total words uttered, number of lexical words that line up with hand gestures, number of function words that line up with hand

gestures, total words that line up with hand gestures, and the percentages of either lexical or function words over total words that line up with hand gestures. The first step was to collect the total words uttered for each video and come up with a transcript. Since the videos are relatively short, for accessible data collection purposes, the process was basically done manually. There was assisting software that made the process easier to do by ear. The videos were first downloaded in mp3 format using 'YouTube to MP3' public online applications and then the mp3 files were imported into a program called Sony Acid Music Studio. The music program read the files and gave a visual representation of the waveforms in each video, which made it easier to stop and know where each phrase ended and began when counting it by ear. The program also allowed for the mp3 files of the videos to be slowed down so that the sound playback would not be faster than the counting.

In regards to counting the lexical or function words uttered to the hand gestures, first, a definition of hand gesture needs to be implemented. Hand gestures in all speeches seem to have a huge variability in how lively they are and where they go. For this specific study, a hand gesture as used for emphasis is defined to be the brisk iteration of the hand in any direction from a very distinguishable stable position. This reference to a hand gesture for emphasis is adapted from the idea of how music conductors use their hands. Conductors define the tactus beat of every measure with a brisk movement of the hand in a specific direction; Whether it'd be down, up, left, or right. The same phenomenon seems to occur when looking at speakers when they move their hands, they have preparatory movement to place their hands but what stands out is the relatively brisk motion that in this study is observed for measure of physical emphasis.

Table 1
Lexical Word Line Up List for Neil Video

Privatization	Unlike	History	Combine	Danger	Moon
Space	Newt	Civilization	Create	Safe	Swallow
Play	Gingrich	Lead	Capital	Winds	Saliva
Future	Said	Large	Market	Established	Caught
Ambitions	President	Expensive	Evaluation	Enterprise	Simple
Space	Debate	Projects	Funded	Investor	Questions
Delusional	Work	Unknown	Governments	Payback	Past
Thinking	Enterprise	Risk	Maps	Private	Done
Privatization	Unlike	History	Combine	Danger	Moon
Space	Newt	Civilization	Create	Safe	Swallow

* Table 1 demonstrates all the observed hand gesture line ups to lexical words from the Neil video. For this video, there were 59 lexical words that lined up with hand gestures which also constitutes for 72.84% of total words that lined up. The total count of lexical words from the sample is 209, and the lexical words that aligned with hand gestures over total lexical words is 28.23%.

Table 2
Unction Words Line up List with Hard Gestures

There	Never	Then	Not	Were	For
For	With	Here	Not	How	Rather
What	Never	We	Was	It	
No	These	So	Get	That	

* Table 2 demonstrates all the observed hand gesture line ups to function words from the Neil video. The table shows that the video contained 22 function words lining up with hand gestures which also constitutes for 27.16% of total words that line up. Such percentage is explicative the significant majority that lexical words had in lining up with hand gestures. These numbers are contrasted to the total of 216 function words in the video, which the percent of function words emphasized over total function words is 10.19%.

Figure 1
Word Frequencies for Neil Video
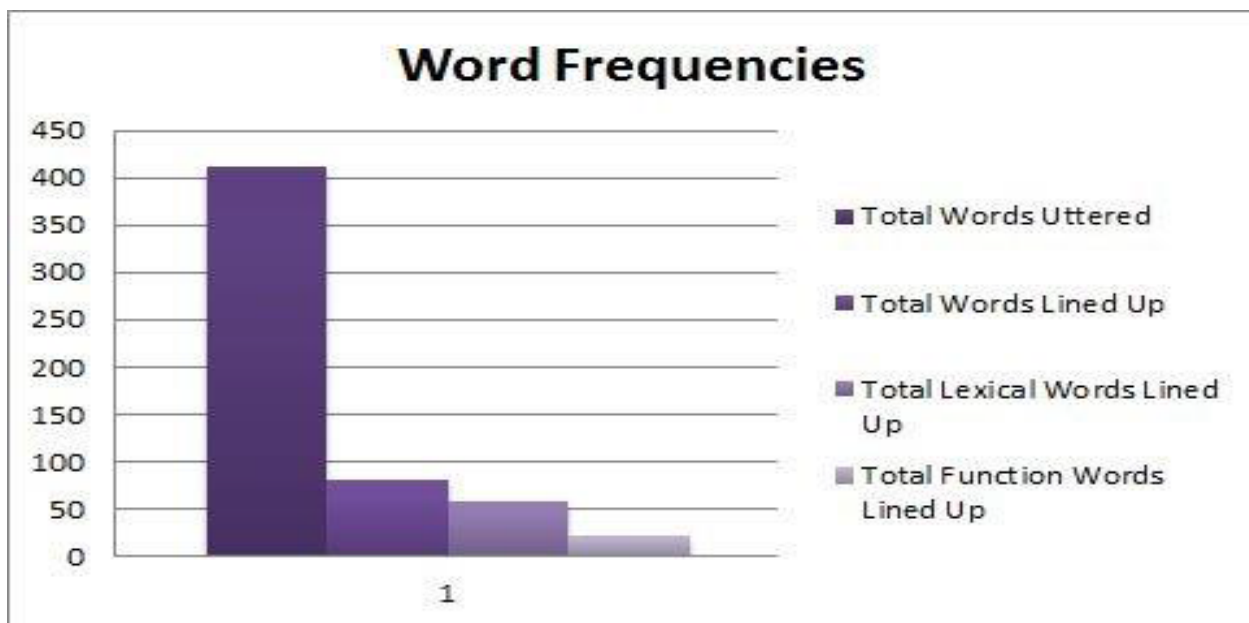
* In Figure 1, the frequencies relative to total words uttered were illustrated. The frequency for total words lined up in relevance to total words uttered shows a significant difference. Such

difference is to be expected and rewards the notion of hand gestures used as emphasis. This is because if a much greater number of words lined up to hand gestures, then it would dilute its use for emphasis. The figure also portrays the significant gap between the majority of lexical words lined up to the minority group of function word line ups.

Table 3
Lexical Word Line Up for Bill Video

Interrupt	Industry	Work	Twenty-Six	Antartic
Let	PG&E	Everything	U.S	Said
Particular	Economy	Wrong	Goals	Change
Skeptics	Issue	Policy	Going	Three-Hundred
Hundreds	Actions	Reaching	Impact	Four-Hundred
Interrupt	Industry	Work	Twenty-Six	Antarctic
Let	PG&E	Everything	U.S	Said

Table 4
Function Word Line Up for Bill

What	Will	Did	That's
Is	What	Not	All
Be	And	To	Really
Not	Will	You	Really

* Just like Tables 1 & 2, the tables above list out the function and lexical word lineups with hand gestures but for the Bill video. The frequency of lexical word line up to hand gestures is 31 and constitutes 65.96% of the total words that line up with hand gestures. The total number of lexical words in the video is 239 and the percent of lexical words lined up over total lexical words is 12.97%. On the other hand, Table 4 lists out 16 function words that lined up with hand gestures and also compose 34.04% of the total words that line up. The total number of function words for this video is 222 and the total percentage of function words that line up over total function words is 7.2%. For this video, the lexical words have the majority again but the percentage their majority is not as high as in the Neil video.

Figure 2

Word Frequencies for Bill video
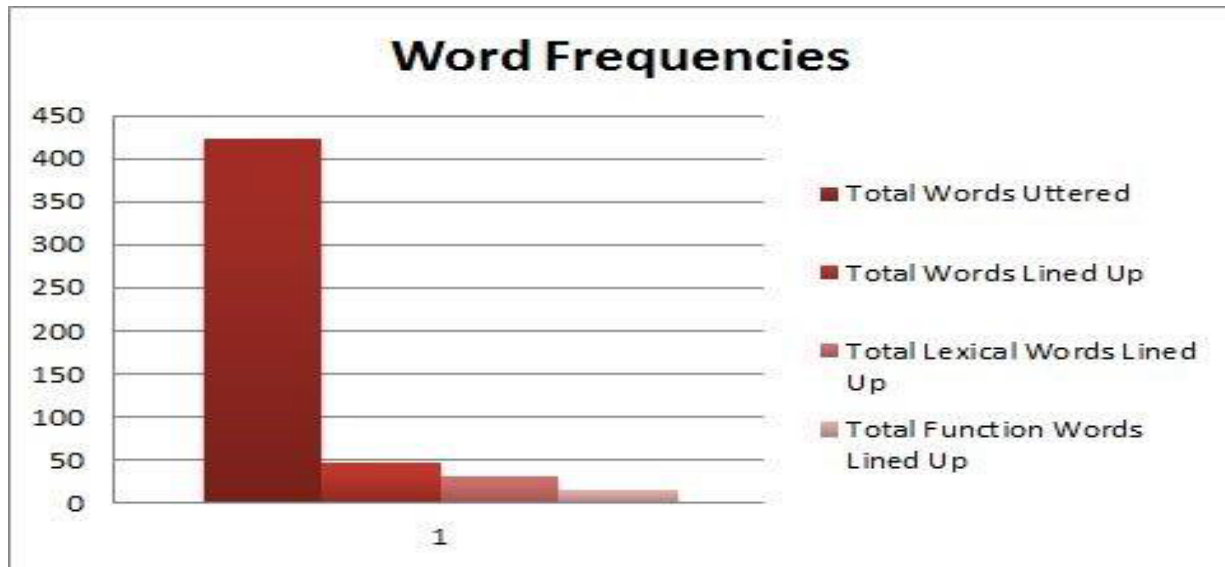
*Figure 2 presents the same type of information as in Figure 1 but for the Bill video. From the figure alone, it is possible to see that the total words line up for the Bill video is much lesser than in the Neil video with the Bill video only having 47 total words lining up and the Neil video having 81 total words lining up. The frequency for lexical words lined up still takes a significant lead in comparison to the frequency of function word line ups, which is to be expected, but the frequency of function word line ups to hand gesture is still surprisingly high in both videos.

When speaking we don't just gesture with our hands, we are also found to use various facial features to gesture in correspondence to our speech. There is evidence that our gestures are actually co-produced with our speech and serve as a physical representation of the information that is being presented. (Kita et al, 2003) If this study would have looked at all perceivable gestures and not just hand gestures, then the number of words lining up would be much higher for both lexical and function words. In fact, there were many times in which other gesture types aligned with words in the same way hand gestures were doing. Although they were not counted, the rest of the other gestures were not used excessively but rather only when it made sense to highlight a word.

Although both videos were different in nature e.g., the Neil video was a unidirectional speech sample and the Bill video was a interactive speech sample, they both aligned with the expectation of having the frequencies of lexical word line ups to hand gestures being the majority in total words lined up. What was more surprising was the percentage of function words lining up in both videos, because although they were the minority, they still had a relatively high count for being expected to not have any important correlation with hand gesture emphasis in speech.

Lexical words in English prosody always appear in stressed form as opposed to function words which are either found to be in a weak stressless form or strong stressed form. (Selkirk,

1996) Words that more constantly receive stress can be expected to be better candidates for emphasis than words that are highly variable on their prosodic stress because these lexical words are standing out in the prosody more often than function words. In a study by David Temperley, he investigates speech corpora and breaks down the prosody in those samples with contextual stress patterning, lexical stress patterning, and interlexical stress patterning. Lexical and inter lexical stress patterning gives data for the expected observance of more stress attributed to words with higher semantic weight which in turn provide higher stress regularity. On the other hand, contextual stress patterning gives data for the instances in which function words are given their strong stressed forms that seem to decrease stress regularity. (Temperley, 2009) This is definitely an example that can be used to explicate why a good amount of function words may be used for emphasis with hand gestures. With reference to Table 2 and Table 4, function words such as 'never','really', and 'rather' can be seen to carry more stress and semantic weight than function words that did not make the list such as 'the' or 'of', which in turn could explain why some function words with higher semantic weight and therefore stress, would more likely be used for emphasis with hand gestures. So, more semantic weight=more stress=more likely to be used for emphasis? This is not necessarily true because there were many instances in which words with higher semantic weight were not emphasized with hand gestures or any other type of gesture. On the other hand, this does not mean that semantics cannot be a predictor for gestural emphasis.

A second look at the transcript extracted from the videos and it is clearly visible that there is another pattern. The majority of function words that are emphasized with hand gestures occur either directly adjacent or within close proximity to lexical words that are being emphasized with hand gestures. What this means for this investigation is that lexical words or words with a significant semantic weight are prime candidates for emphasis, and such semantic weight is an attractor for emphasis which also encapsulates other neighboring function words. For example, if a speaker chooses to emphasize a certain point in their argument, then it can be expected that the words aligning with gestural emphasis have high semantic weight and it attracts gestural emphasis towards other neighboring words with significantly lower semantic weight. There are some examples from the videos in which function words line up with gestural emphasis in an isolated nature. Part of this phenomenon can be explained by the fact that the videos were limited in perspective, because the camera would shift focus away from the hands and during that blind period for the hands there could have been a hand gesture lining up with an adjacent lexical word. Another explanation for this phenomenon can also be traced back to semantics; at times were the hands are clearly visible and only line up with a function word and no neighbors, those function words tend to be crucial for the presentation of the speaker. These function words are words such as 'rather' or 'not' and their primary job is to change the tone or direction of an argument which may be crucial for the speaker to point out.

With this analysis, there is an evident new method to find certain gestural emphasis more predictable. Certain variables that may skew the predictability of gestural emphasis may be the different stylistic directions taken in a presentation by a speaker and their biological tendencies such as stuttering or pausing frequently. Besides these factors, there is evidence for the predictability of gestural emphasis in speeches whether it can be said that such coordination is highly predictable or not so predictable. For these two video samples, the same patterns occur and gestural emphasis is for the most part distributed in all the sections of the speech. This observation definitely opens up the path to ask whether such distribution and predictability can demonstrate isochrony with gestural emphasis in speech or a grounded pattern of regularity that could imply a rhythmic tendency towards such presentations.

A study looking at German speakers and beat gestures has demonstrated evidence for gesturing and speech as an integrated system working in a rhythm dependent setting. The beat gestures studied would highlight noun phrases and make it easier for the subjects to comprehend more syntactically complex sentences because over time they would expect these beat gestures to show up regularly and give the speech a more predictable pattern. (Holle et al, 2012) That study is a prime proxy example of how gestural emphasis may play a role into information predictability. If a rhythmic beat gesture can be used to help listeners digest more syntactically complex sentences then it is valid to say that gestural emphasis can also have the same effect. Despite the stress irregularity that function words bring, it can be concluded that a master of speech presentations can will tactfully play with surprise and expectations. A rhythmic approach to a presentation may help digest more complicated parts, but the dynamic stress irregularity brought upon by function words may be a needed tool in surprise that can keep the audience at the edge of their seats.

This study has several limitations to the claim at hand such as only having 2 videos for the sample size. The videos also have parts in which the focus is moved away from the hands and the viewer misses out on a number of hand gestures. Therefore, this data only took account of hand gestures that were immediately clear and even with conservative numbers there exists evidence for the predictability of gestural emphasis and its possible use in reference to rhythm. The analysis of a larger sample pool would be beneficial in having stronger numbers that could generalize the extent of the predictability of gestural emphasis in speech. Even though this study may not solve the mysteries of speech and rhythm as belonging to an integrated cognitive system, it does provide insight and enough evidence to continue the dialogue of the idea of rhythm being implemented in our everyday speech deliberations in a limited way.

References

[1] Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*(13), 1176-1182.

[2] Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905-922.

[3] Granström, B., House, D., & Karlsson, I. (Eds.). (2013). *Multimodality in language and speech systems* (Vol. 19). Springer Science & Business Media.

[4] Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech.*Frontiers in psychology*, *3*.

[5] Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*(1), 16-32.

[6] Selkirk, E. (1996). The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, *187*, 214.

[7] Temperley, D. (2009). Distributional stress regularity: A corpus study. *Journal of psycholinguistic research*, *38*(1), 75-92.

[8] Winter, B., Perlman, M., & Matlock, T. (2013). Using space to talk and gesture about numbers:Evidence from the TV News Archive. *Gesture*, *13*(3), 377-408.

Biography

I am a fourth year Cognitive Science major at UC Merced who is currently applying for graduate school in hopes to get into a Ph.D program for Cognitive Information Sciences. My research experience with UC Merced's Psycho-Neuro-Endocrinology Lab, Theory of Mind Lab, Dr.Jeff Yoshimi, Dr.Stephanie Shih, and Dr.Chris Kello at the Cognitive Mechanics Lab has allowed me to grow my skills outside the classroom and prepare for research after graduation. My passion as a musician and involvement as a Cognitive Science student is something I seek to integrate, and one day hope to be doing exciting research at the doctoral level like the many professors who have motivated me by being great role models."