INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 666-2900 • FAX (510) 666-2956

Acoustic Stress and Topic Detection in American English Spoken Sentences

Rosaria Silipo, Fabio Crestani

TR-00-005

March 2000

Abstract

The relationship between acoustic stress and information content of words is investigated. On one side, the average acoustic stress is measured for each word throughout each utterance. On the other side an Information Retrieval (IR) index, based on the words frequency throughout the particular spoken sentence and throughout the collection of analyzed spoken sentences, is calculated. The scatter plots of the two measures (average acoustic stress on the y-axis and IR index on the x-axis) show higher values of average acoustic stress with the increasing of the information measure of the word in the majority of the analyzed utterances. A statistically more valid proof of such a relationship is derived from the histogram of the words with high average acoustic stress vs. the IR index. This confirms that a word with high average acoustic stress has also a high value of the IR index.

1 Introduction

Prosodic stress is sometimes intended as lexical stress, as in [1], and sometimes as acoustic stress, that is as an acoustic enhancement of some parts of the utterance, as in [2] and [3]. This acoustic enhancement is often referred to as prominence [3] and, together with many other prosodic features, represents a means for the speaker to enhance and correctly transfer information, to communicate emotions and so on.

The two stresses, lexical and acoustic, are somehow related, because often the acoustic enhancement of a word is centered around its lexically stressed syllables. However there is not a one to one correspondence. Very often speakers stress an entire word or sequence of words, which means that all syllables of the word are acoustically enhanced, even the lexically unstressed ones. Other times unimportant words, like articles or conjunctions, are acoustically not perceived as prominent even though in the dictionary they exhibit a lexical stress.

It is by now known that there is a relationship between acoustic stress – or prominence – and information spotting. This phenomenon has been mainly investigated inside dialogues, where stressed words occur to introduce new concepts [4]. Not much research has been developed so far to transfer this investigation about the introduction of new concepts and the occurrence of acoustically stressed words from dialogues to monologues. It is very reasonable though that speakers use the same psycho-acoustic effects when talking for longer time.

If it is possible to establish a correspondence between acoustically stressed words and information, an automatic detection of acoustic stress would also provide a measure of the information associated with each word of the utterance. This would help to define some "acoustic" keywords, that would be related with the informative keywords and could be used for indexing spoken documents and IR purposes.

So far, in the Information Retrieval (IR) literature, the detection of keywords in texts is performed by means of heuristic statistic-based indexes (see for example [5, 6]). Those words that are not so frequent throughout the collection of documents but slightly more frequent inside a particular document can be reasonably considered representative of the document's contents. This theory goes back to Zips and Luhn [16]. That is, such words can be used as keywords to summarize and retrieve this particular document. Given a collection of spoken documents, if it is true that acoustic stress enhances systematically the informative words of each monologue, acoustically stressed words and IR keywords should, at least partially, overlap. In addition, acoustic stress could provide complementary information to the IR indexes and help the identification of more accurate keywords for the characterization of spoken documents.

In this work, we try to investigate the relationship (if any) between acoustic stress and informative content of the words. For this reason we focus on those words that are acoustically stressed, because they are information related. In general words in a spoken sentence can be acoustically stressed because they are associated with information or because they occur inside an emotional context, carry contrastive stress or for any other reason. The goal is to separate the first group of words from all the others and to see whether they can be used as keywords to characterize the whole spoken sentence.

To isolate words that are acoustically stressed because information related we average the acoustic stress across the number of occurrences of each word in the utterance. Words that are information related are going to be systematically stressed, that is for most occurrences, while words that are only

occasionally stressed, because they happen to be in an emotional context or they carry contrastive stress, will be stressed only for a little fraction of the occurrences. The average operator will penalize words that are only occasionally stressed and therefore not information related.

These words that are stressed in average throughout the sentence are also evaluated by using an IR index to see whether they can be considered as keywords to characterize the monologue. In this work, words with low average acoustic stress show a low value of the IR index and an increasing average stress usually goes together with an increasing value of the IR index.

2 Prosodic Stress

Prosodic stress is an integral component of spoken language [2], particularly for languages such as English that so heavily depend on this parameter for lexical, syntactic, and semantic disambiguation. The relationship between prosodic stress and semantic disambiguation represents the scope of this study.

2.1 What is Acoustic Stress?

We refer here to acoustic stress rather than to syntactical stress, that is to the acoustic perception of a syllable or group of syllables as enhanced with respect to the others. This kind of acoustic stress is often referred to as *prominence* [3, 2, 7]. The two stresses, acoustic and syntactic, might not be unrelated since the grammatical structure should give the guidelines for enhancing one or another part of the sentence. The hypothesis to be investigated is whether any connection exists between the emphasis given to a word or a group of words and the content of the spoken sentence.

Experimental and descriptive studies [7, 8, 9] indicate that acoustic stress is mainly based on a complex constellation of information pertaining to the duration, amplitude, and fundamental frequency (pitch) associated with syllabic sequences within an utterance. Because a large part of prosodic stress information is carried by the vocalic nucleus [10, 7], the acoustic stress is often associated with the vocalic part of the syllable.

In figure 1.a, some examples of stressed syllables in a spoken sentence are shown, where five syllables were marked as stressed by two trained linguists. Figure 1.b) shows the corresponding fundamental frequency plot. The last four stressed vocalic nuclei are characterized by a longer duration and the two dyphtongues "ay" also by a higher amplitude, while the first stressed vocalic nucleus is characterized by a high constant value of the fundamental frequency. The traditional perspective attributes the perception of prosodic stress to pitch height or pitch variation [3] and only up to a lower extent to duration and amplitude. Recent studies, however, focusing on spontaneous American English speech, found pitch less effective than duration and amplitude [9].

In the sentence reported in figure 1, the message is contained in the words "riding", "bikes", "kids" and up to a lower extent in "like" and "do". All these words, except for "like", are labelled as acoustically stressed by two linguists.

2.2 The Annotated Corpus

In order to provide a basis for the proposed investigation, the prosodic stress of a portion of the American English component of the OGI Stories Corpus [11] was manually marked by two trained linguists.

Many levels of prosodic stress are usually reported in the literature. However, the concordance among linguists becomes lower and lower with a higher number of stress levels [12]. It can be safely

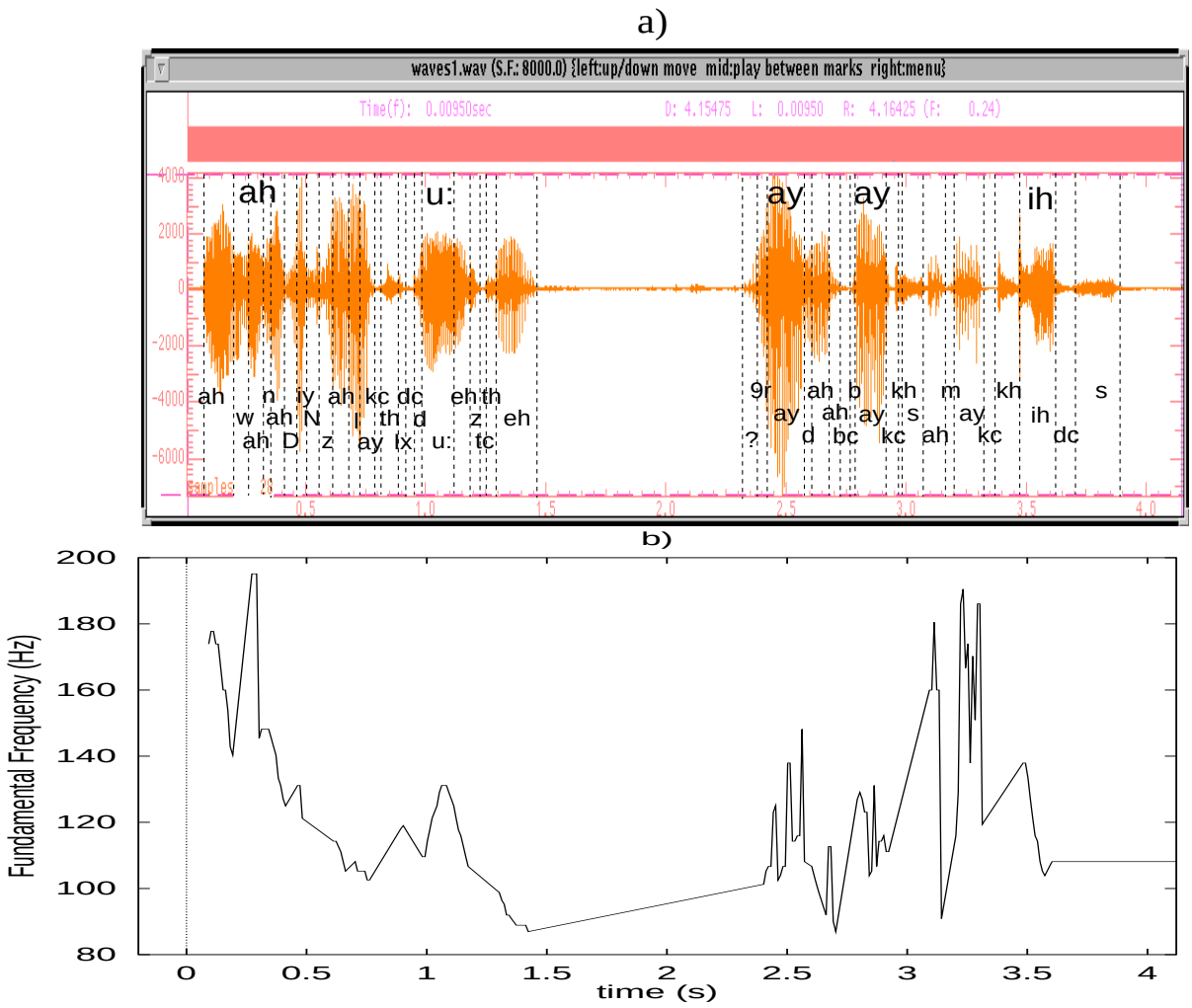

Figure 1: a) Examples of stressed vocalic nuclei in"uh, one of the things I like to do is to .pau ride with bikes with my kids". On the bottom there is the phonetic segmentation of the spoken sentence and on the top the stressed syllables. b) Fundamental frequency plot for the spoken sentence in a).

Table 1: Number of files from the OGI Stories database labeled by each transcriber.

voices	first transcriber	second transcriber	both
men	49	38	5
women	34	13	5
total	83	51	10

Table 2: Percentages of stressed syllables according to each transcriber.

	S	S+	S-	N
transcriber # 1	32%	15%	17%	68%
transcriber # 2	23%	18%	5%	77%

assumed that only three levels of prosodic stress can be reliably detected by trained linguists: a primary stress (S+), an absence of stress (N) and an intermediate category of syllables with weak stress (S-).

The OGI corpus contains 50-60 second files about a variety of subjects. A phonetic transcription of the files is also supplied. Two different subsets of files, including only spontaneous American English speech, are extracted from the database and separately annotated in terms of prosodically stressed (S+ or S-) syllabic utterances. The first subset, annotated by transcriber # 1, includes 83 files, with 49 men and 34 women voices. The second subset, annotated by transcriber # 2, contains 51 files, with 38 men and 13 women voices. 10 files, 5 men and 5 women voices, are common to both subsets of files and are then annotated by both transcribers (Table 1).

2.3 Statistics of the Two Annotated Subsets

In order to describe the acoustic stress pattern marked by the two transcribers, the corresponding statistical distribution and interval histograms across the spoken sentences are evaluated.

Table 2 reports the percentages of primary (S+), minor stressed (S-) and unstressed (N) syllables across the two data subsets annotated by the two transcribers. In the first columns the percentage of stressed syllables (S including both S+ and S-) is reported. Both transcribers perceive circa 20-30% of the syllables as stressed, which agrees with previous studies [9]. However, transcriber # 1 detects a higher percentage of stressed syllables than transcriber # 2, while transcriber # 2 seems to be more biased towards marking primary stress.

From table 2 the general distribution of stressed and unstressed syllables is made clear. However, the evolution pattern of primary and minor stresses across a spoken sentence can not be derived. In order to explore such time trend, interval histograms are constructed for different combinations of stressed and unstressed syllables.

In figures 2.a and 2.b we refer to stressed syllables (S, including both S+ and S-) that are separated by a number x of unstressed (N) syllables, such as for example the trend S N N N S that corresponds to $x = 3$. The number of unstressed (N) syllables inbetween two stressed syllables (S) is reported on the x-axis and the corresponding proportion of stressed syllables on the y-axis. Transcriber # 1 grants stress to syllables that are consecutive or far apart just one unstressed syllable in the 60% of the cases. Transcriber # 2 considers circa the 60% of the stressed syllables to occur

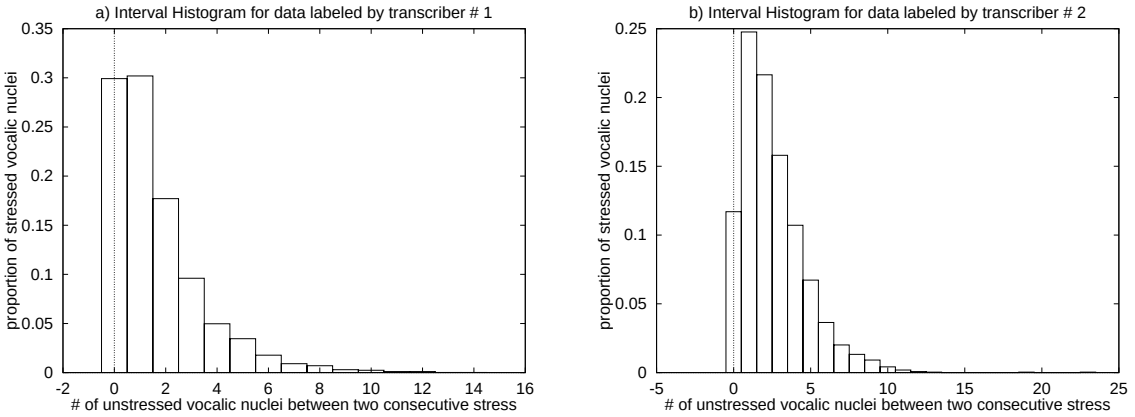

Figure 2: Interval histograms for stressed vs. unstressed syllables a) for the first and b) for the second transcriber's data.

after one, two or three unstressed syllables and only more rarely (slightly above 10%) one after the other. In general, however, there is a high probability that stress occurs again within four or five unstressed syllables. The maximum interval of unstressed syllables between two stresses is around 20 for transcriber # 2 and around 14 for transcriber # 1. This certainly derives from the highest frequency of stress perceived by transcriber # 1 with respect to transcriber # 2.

Let us analyze the stress pattern in more detail. In figures 3.a and 3.b we refer to primary stressed (S+) syllables that are separated by a number x of unstressed syllables (N). The number of unstressed (N) syllables inbetween two primary stresses is reported on the x-axis and the corresponding proportion of primary stresses on the y-axis. According to transcriber # 1, primary stress does not occur regularly. Indeed around 75% of primary stress occurs immediately, one, two or three unstressed syllables after another primary stress. However, the remaining 25% of primary stressed syllables occurs after a higher number of unstressed syllables, between five and sixty-four. This happens in particular cases, when the person's voice does not use very much stress during his/her speech. Transcriber # 2 shows an interval histogram very similar to the one in figure 2.b. Indeed, as we can see from table 2, transcriber # 2 labels almost all stresses as primary stress and consequently the two interval histograms are very similar.

In figures 4.a and 4.b we refer to minor stressed (S-) syllables that are separated by a number x of unstressed (N) syllables. The number of unstressed syllables inbetween the two minor stresses is reported on the x-axis and the corresponding proportion of minor stress on the y-axis. Transcriber # 1 divides stressed syllables in almost the same proportion of primary and secondary stress (Tab. 2). The two categories though are not equally distributed across the files. Some people make use of stress in a very clear way, while some other people do not rely on stress as much to be understood. For this reason, the interval histogram of minor stressed vs. unstressed syllables of transcriber # 1 is close to the one in figure 3.a, because the same amount of primary and minor stress results to be more concentrated in some files rather than in others. The interval histogram of minor stressed vs. unstressed syllables for the second transcriber is spread over the [0:20] range of unstressed syllables inbetween two minor stresses. This is likely due to the low number of minor stresses detected by transcriber # 2.

Finally, let us examine the relationship between primary and minor stress only, without considering unstressed syllables. In figures 5.a and 5.b the number of minor stressed (S-) syllables inbetween

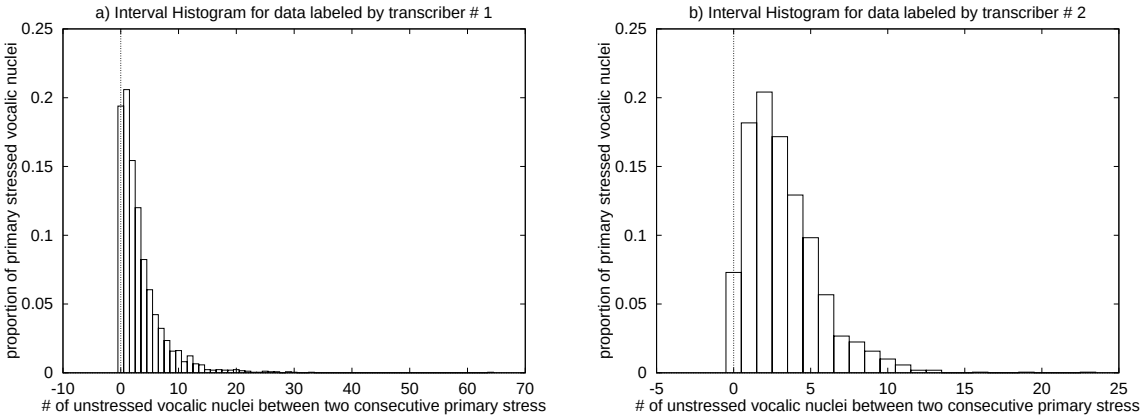

Figure 3: Interval histograms for primary vs. unstressed syllables a) for the first and b) for the second transcriber's data.

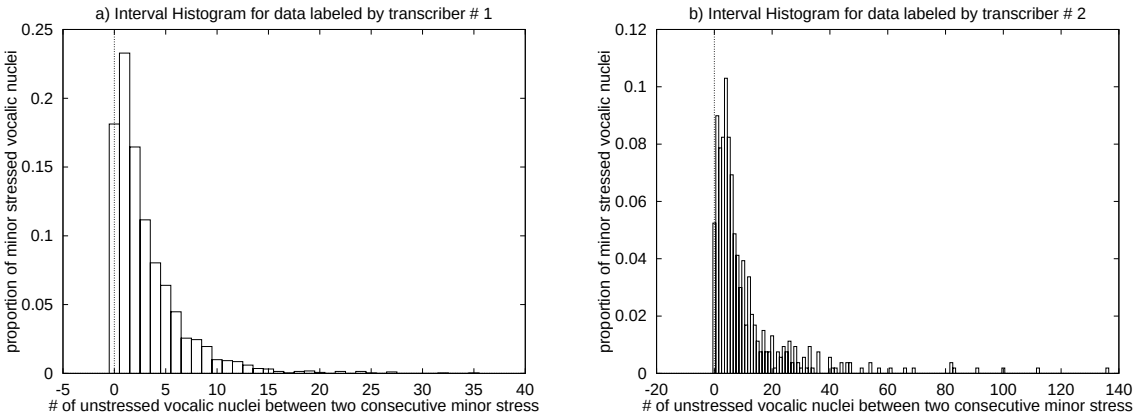

Figure 4: Interval histograms for minor vs. unstressed syllables a) for the first and b) for the second transcriber's data.

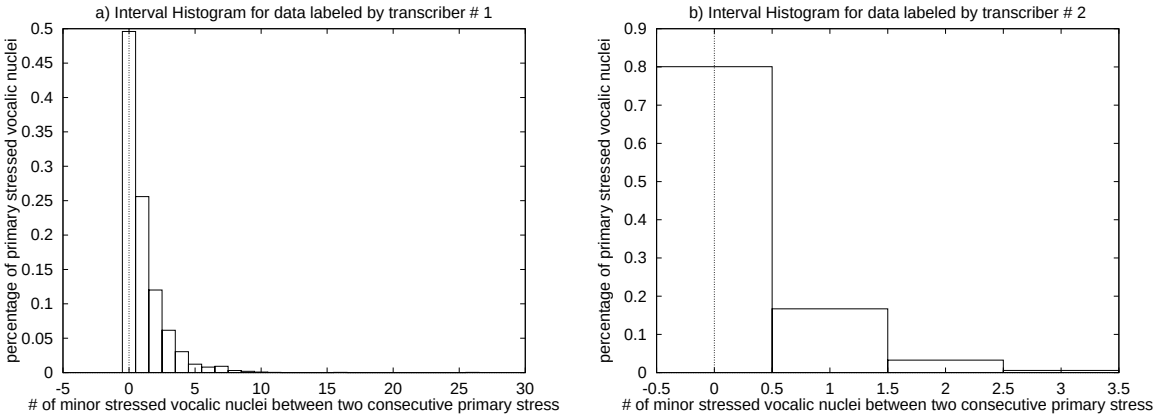

Figure 5: Interval histograms for primary vs. minor stressed syllables a) for the first and b) for the second transcriber's data.

two consecutive primary stresses (S+) is reported on the x-axis and the corresponding proportion of primary stresses on the y-axis. Both figures show that a very high number of primary stresses (50% according transcriber # 1 and 80% according to transcriber # 2) occur immediately after another primary stress, without any minor stress in the middle. However, from figures 2.a, 3.a and 4.a the percentage of primary stresses following another stress is very high. Thus, primary stress and secondary stress are often combined together, corresponding to those words that the speaker wants to emphasize.

2.4 Transcribers' Agreement

The agreement between the two transcribers on the common files between the two OGI data subsets is reported in Table 3.

The human perception of different kinds of stress (like S+ and S-) is usually very controversial. For this reason, only two levels of stress are considered to measure the agreement between the two transcribers: stressed (S), including primary and minor stress (S+ and S-), and unstressed (N) syllables. Accordingly a stress labeled as S+ or S- by one transcriber is considered in agreement if it was labeled as either S+ or S- by the other transcriber.

The first three columns of Table 3 refer to the agreement percentage of transcriber # 1 vs. transcriber # 2, considering only men voice files (M), only women voice files (W) and both together (W+M). The second three columns refer to the agreement percentage of transcriber # 2 vs. transcriber # 1. The last three columns refer to the average agreement percentages. The two transcribers roughly agree in recognizing unstressed syllables (N: 84-93%) and primary stress (S+: 90-78%). Much more disagreement exists in recognizing minor stresses (S-: 67-57%). The high agreement percentages about unstressed syllables (N: 84-93%) and primary stressed syllables (S+: 90-78%) show the reliability of the prosodic stress hand-labelling process.

2.5 Quantifying the word's emphasis

The goal of this study is to characterize the information associated with each word by means of acoustic stress and of IR statistical indexes. The acoustic stress characterization then has to move

	Transcr. # 1 vs. # 2			Transcr. # 2 vs. # 1			Average agreement		
	% agreement			% agreement			% agreement		
	S+	S-	N	S+	S-	N	S+	S-	N
W+M	90	67	84	78	57	93	84.0	62.0	88.5
M	93	76	84	81	58	94	87.0	67.0	89.0
W	87	46	85	74	56	92	80.5	51.0	88.5

Table 3: In the first three columns, agreement of transcriber # 1 vs. transcriber # 2 and, in the second three columns, agreement of transcriber # 2 vs. transcriber # 1 and in the last three columns the average agreement of the two transcribers on all the common files (W+M), only the male speakers common files (M) and only the female speakers common files (W). S+ primary, S- minor stressed, N unstressed syllables.

from a syllabic to a word perception. For each word the total sum of syllabic stress is evaluated attributing +1.0 to primary stress and +0.5 to minor stress. Most English words are monosyllabic, that is the occurrence of more than one stress per word is a quite rare phenomenon.

We divide the occurrences of acoustic stress in two main classes. On one side the words that carry acoustic stress because they are information related and on the other side all the other words that are stressed because of any other reason (emotional context, contrastive stress and so on). The idea is that if a word is information related, it will be stressed, and if it is repeated more than once, it will be stressed in many of its occurrences. Words that are stressed but not information related, are only occasionally stressed, that is this particular occurrence might be stressed but in general the word can be found again as unstressed. In order to characterize acoustically stressed words that are information related, an *average acoustic stress* is defined across all occurrences of each term within an utterance. The average operator penalizes words that are only occasionally stressed, due to the particular context and/or to a particular stress pattern in which they occur. Keywords on the opposite should benefit from one or more occurrences acoustically stressed throughout the sentence.

Based on the previous considerations, the average stress is defined across all the instances of the same word within each utterance, as described in table 4. Thus the average amount of stress can be used to characterize the importance of each word of the sentence.

Many words in English are monosyllabic and consequently their total amount of stress per instance can be at most 1.0. The average stress described in table 4 is adopted as a measure of the information associated with the word, under the hypothesis that a keyword for the sentence is stressed for most of its occurrences. Thus the more important the word is the higher its average stress is. However, in the OGI Corpus sentences keywords are either pronounced only once or are always repeated with same amount of stress. Therefore, most of the words exhibit three possible values for the average stress: 1.0 if the word is characterized by primary stress, 0.5 if it carries minor stress, 0.0 if it is unstressed in all the occurrences in the sentence. The majority of the words have an average stress as 0.0, that is they do not exhibit any stress in any of their instances.

In figures 6.a and 6.b the histograms of the words vs. their average stress are reported for transcriber # 1 and # 2. As we can see the highest percentage of words has an average stress as 0.0. All other words are mostly located at the 0.5 and 1.0 average stress points. Only a small groups of words present an average stress outside of these three main lines. For the second transcriber the lower number of words with minor stress generates a smaller bar at average stress 0.5.

Instances and stress per instance for the word "program" labeled by transcriber # 1		
instance #	instance of the word "program"	stress per instance
1 2	pro(S+)gra(S-)m progra(S+)m	1.5 1.0
total # of instances = 2	Total stress for word "program" = 2.5	
Average stress $= \frac{\text{total stress}}{\text{total \# of instances}} = 1.25$		

Table 4: The syllabic stresses on each word are summed together (1.0 for primary stress S+, 0.5 for minor stress S-) to obtain the amount of acoustic stress characterizing this word. The total stress is calculated by summation across the number of instances of the word inside the sentence. The average stress is obtained by dividing the total amount of stress for the number of instances of the word in the sentence. Here the example of the word "program" in file en054stb.phn labeled by transcriber # 1 is reported.

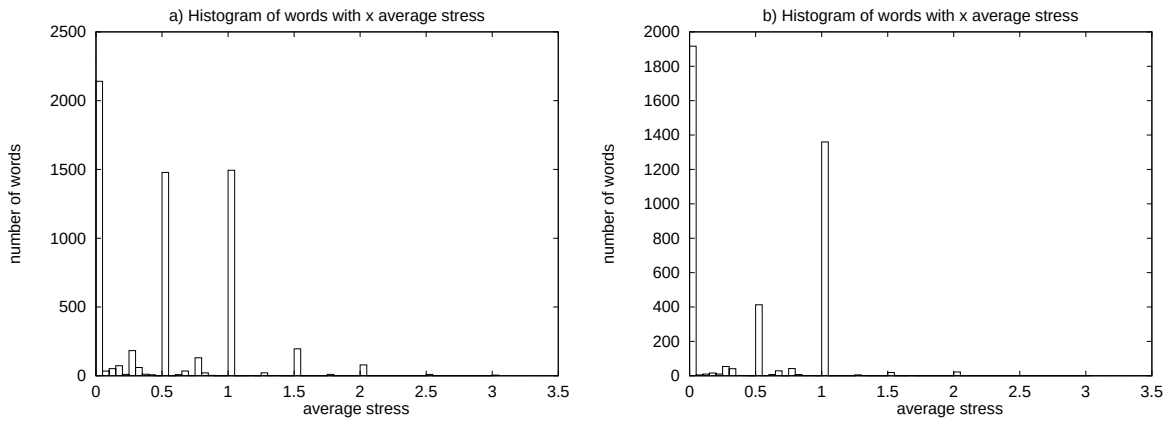

Figure 6: Histograms of the words vs. their average stress inside each file a) for the first and b) for the second transcriber's data.

3 The Information Retrieval Index

Topic detection can be identified as a subtopic of indexing. Indexing, that is the extraction and weighting of *index terms* from the text of a document in order to represent the document informative content, has been studied for long time in IR (see for example [18, 21]).

The most well established techniques rely on exploiting statistical information about term occurrences. According to the most widely accepted indexing technique [5], indexing is divided into the following steps:

1. term extraction;
2. stop words removal;
3. stemming and conflation;
4. index term weighting.

Term extraction is the first phase which consists in breaking the full text of the document into words[1]. This operation reflects the normal file organization used in IR, which is the inverted file, using a term list with linked document identifiers, plus counting data, and pointers to the actual texts.

Index terms are normally content words, so during the *stop words removal* step stop words (e.g. prepositions, articles, conjunctions) are identified via a stop word list and removed from the inverted file. In the experiments reported in this paper we did not remove stop words, since we wanted to consider all terms present in document. Normally, a stop word list like the one reported in [21, pp.18] would have been used.

Index terms are also generally word stems or roots rather than full words, since this means that matches are not missed because of trivial word variations, such as with singular/plural forms, or verb tenses. *Stemming and conflation* are usually carried out using a standard stemming algorithm and suffix list. A number of algorithms for such use exist (see for example [15]), and in the experimentation reported here we used the Porter's algorithm [17], one of the most widely accepted ones.

After the above three phases, indexing could stop. We would then have a document representation which is "binary", that is based only on the presence or absence of terms from the text of the document. This approach was used by a number of systems in the past and is still in use for Boolean IR systems. However document indexing may be considerably improved by giving a weight to each term-document combination.

The *index term weighting* phase is the most complex and time consuming phase of indexing since it involves carrying out a large number of computations based on the counting of the frequency of occurrence of terms in each document and in the collection as a whole.

The idea behind index term weighting is "discrimination": a good index term should be able to characterize some part of the document informative content, distinguishing that particular document from all others. To be a good index term for a document the term should be frequently occurring in the document, taking into account the document length, but not so frequently occurring in the whole collection. Therefore, it is easy to see that there are three different sources of weighting data:

1. term frequency;

[1] We are not concerned here with multi-word string or phrase indexing.

2. document length;

3. term collection frequency.

Among the many different ways of defining index term weights based on the above sources, we chose the (arguably) most theoretically sound and widely experimented schema, the one used by the Okapi experimental IR system and described in [5, 6].

In the Okapi schema, *term frequency* is defined as the term's within document frequency of occurrence. In other words, given a term t_i and a document d_j, $tf(i,j)$ is the number of times t_i occurs in d_j.

The term frequency cannot be used as it stands as a weighting factor, since it is influenced by the length of the document. This is the reason why the *document length* needs to be taken into consideration. Given a document d_j, the document length $dl(j)$ is the total number of term occurrences in d_j. However, a better weighting factor is the normalized document length $ndl(j)$, defined as:

$$ndl(j) = \frac{dl(j)}{(\sum_{k=1}^{N} dl(k))/N}$$

where N is the number of documents in the collection and therefore $(\sum_{k=1}^{N} dl(k))/N$ is the average document length.

Finally, the third weighting factor, *term collection frequency*, is motivated by Luhn's law [16] on term distribution. The rationale of this factor is that terms that occur in only a few documents are often more valuable discriminators than terms that occur in many. So, given a term t_i, the term collection frequency weight $cfw(i)$ (often known as inverse document frequency, *idf*), is defined as:

$$cfw(i) = \log N - \log n$$

where N is the number of documents in the collection, and n is the number of documents the term t_i occurs in. The use of the logarithm is motivated by the need of smoothing this weighting factor.

After having computed the three sources of weighting evidence, their combination is obtained using the following formula:

$$w(i,j) = cfw(i) \cdot \frac{(K_1 + 1) \cdot tf(i,j)}{K_1 \cdot ((1 - K_2) + K_2 \cdot ndl(j)) + tf(i,j)}$$

where all symbols have already been defined apart from the two tuning constants K_1 and K_2. This formula ensures that the effect of the term frequency weight is not too strong, and for a term occurring only once in a document of average length, the weight is just $cfw(i)$. The constant K_1 modifies the extent of the influence of $tf(i,j)$. Ideally it should be set after systematic trials on the particular collection of documents being used. The constant K_2 modifies the effect of $ndl(j)$. Typically, if $K_2 = 1$ the assumption is that documents are long simply because they are repetitive, while if $K_2 = 0$ the assumption is that documents are long because they are multi-topic. In the experiments reported in this paper we set $K_1 = 1$ and $K_2 = 0.5$, after considering the similarities of the collection of documents being used with some standard IR text collections.

In the rest of this report we will refer to the $w(i,j)$ weighting evidence as *tf_idf* weight, since this is the name most often associated to this evidence in IR literature.

In figures 7.a and 7.b the histograms of the words vs. their *tf_idf* [13] value are reported for both transcribers' data. Many words have a 0.0 *tf_idf* value, that is articles and other unspecific words (i.e. most words would appear in the stop list). All the other words are more or less equally distributed across the remaining *tf_idf* values.

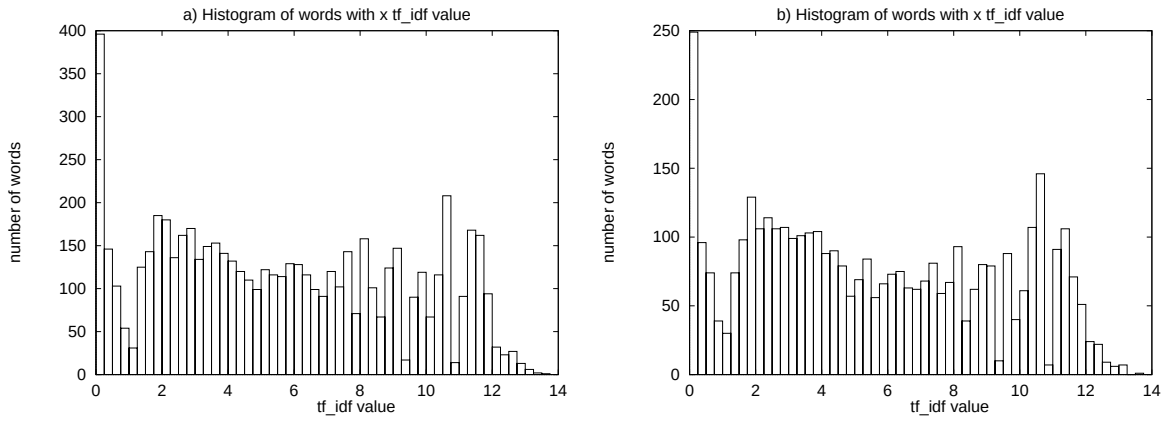

Figure 7: Histograms of the words vs. their tf_idf value inside each file a) for the first and b) for the second transcriber's data.

4 Relationship between the Information Retrieval Index and the Average Acoustic Stress

The goal of this work is to investigate the relationship, if any, between acoustic stress and informative character of a word in a spoken sentence. In particular, we refer here to short monologues as the ones found in the OGI Stories Corpus rather than to dialogue sentences. In this context the informative character of a word becomes its capability of characterizing the content of the monologue, that is of being a *keyword*. If any relationship exists between acoustic stress and keywords, an approach to speech document indexing could be attempted using the acoustic stress as information measure.

In the previous section, an IR index (tf_idf) has been introduced, in order to supply a measure of the capability of each word to characterize the content of the spoken sentence. High values of the IR index are associated with a high informative character of the word.

In section 2.5 a measure of the average acoustic stress is defined for each word throughout the spoken sentence. This measure quantifies how much a given word is stressed in average across all its utterances. Words that do not carry much information are usually repeated very often, like articles, conjunctions and prepositions, and are only rarely and not systematically acoustically stressed, like in the case of contrastive stress or emotional context. On the other hand keywords are usually introduced the first time with a considerable acoustic stress and repeated a number of other times, many of them with acoustic stress. The acoustic stress pattern is then more systematic for these words than for the other non informative words. Calculating the average acoustic stress for each word throughout the utterance allows the rejection of stressed but not information related words and the enhancement of keywords.

In the following sections a visual investigation and a statistical analysis are attempted, in order to visualize and estimate any correlation between the two variables, average acoustic stress and IR index. The visualization is obtained by means of scatter plots and the statistical estimation is performed through the histograms of the words with high average acoustic stress vs. tf_idf values and vice versa.

4.1 Scatter Plots and Keywords

In order to investigate the relationship between acoustic stress and information, in each file of the annotated OGI Corpus each word is represented by a two-dimensional vector that consists of its *tf_idf* value and its average acoustic stress. These points are then mapped on a two-dimensional graphic, being the *tf_idf* index on the y-axis and the average acoustic stress on the x-axis.

If the hypothesis that highly acoustically enhanced words carry most information is true then the scatter plots should show two clear groups of data, one on the bottom left corner, that is for low values of *tf_idf* and average acoustic stress, and one on the top right corner, that is for high values of both *tf_idf* and average acoustic stress. Maybe a third group could be hypothesized, consisting of moderately informative and moderately stressed words. Such third group should then be located inbetween the two other ones.

4.1.1 Speech files labelled by transcriber # 1.

Due to the discrete character of acoustic stress (0, 0.5 and 1.0) the points in the scatter plots are organized along parallel horizontal lines with a distance $\Delta y = 0.5$ between each other. Three main lines can be noticed, one at $y = 0$, one at $y = 0.5$ and one at $y = 1.0$, corresponding to the three discrete values that acoustic stress can take.

The scatter plots of the files labelled by transcriber # 1 generally concentrate the words around a diagonal line with a high concentration of words at the top right and at the bottom left corner. However, the two groups of words are not cleanly separated but a progressive transition from the bottom to the top corner can be observed.

In the appendix the scatter plots of twelve files (from figure 12 to figure 23) annotated by transcriber # 1 are presented. The corresponding uttered sentence is reported in the caption. These files have been selected as the most representative for the material labelled by transcriber # 1. Indeed, all those figures present clearly the diagonal organization of the words along horizontal lines and particularly figures 14, 16, 18, 19, 21 and 22. On the opposite in figures 12, 13 and 15 the diagonal organization of the points is slightly harder to see. Figures 12, 13, 14 and 15 are taken from the set of common files labelled by both transcribers. For these files the stress pattern is also shown in the caption.

The dotted line drawn at $y = 0.6$ marks a hypothetical separation between highly and low average acoustically stressed words. If informative words carry a systematic stress patterns, just by reading the words above this line it should be possible to understand the topic of each monologue described in the figures from 12 to 21. While this is certainly true for figures 14, 16, 17, 18, 19, 20, 21 and 22, this is not so automatic for all the other figures. Generally this depends on two factors: the content of the monologue and the speaker's stress pattern. In some figures the speaker does not have a specific subject but rather jumps from one topic to another trying to filling the time during which he/she has to talk (for example, figures 23 and 15). In other cases, the speaker does not have a very large stress dynamic. In these cases, the stress range is quite small and only a few words present a considerable average acoustic stress (for example, figures 12 and 13).

4.1.2 Speech files labelled by transcriber # 2.

The scatter plots of the files labelled by transcriber # 2, reported in the appendix, are very similar to the ones of the files labelled by transcriber # 1. A representative subset of such scatter plots is shown from figure 24 to figure 35. The diagonal organization of the words shows clearly in these graphics. The only difference with the scatter plots in the previous subsection consists of a lower

number of points on the horizontal line at $y = 0.5$. This reflects the lower number of intermediate stressed syllables perceived by transcriber # 2, as it can be seen in figure 6.b.

By comparing figure 12 with figure 24, we can see that transcriber # 2 transfers many intermediate stresses from transcriber # 1 to primary stresses. In this case then the detection of keywords becomes easier. Indeed the range for the average acoustic stress is reduced and many keywords can now be found as primarily stressed words.

4.2 Statistical Analysis

The previous section showed a number of scatter plots for monologues from both transcribers' data sets, where words with high average acoustic stress seem to be associated with high information content at least according to the adopted IR index. However this scatter plots' pattern is easily detectable for some files and extremely confused for others. This is generally due to the speaker's attitude towards stress and to the monologue's content. Indeed some speakers have a flatter way of speaking than others. They use a very limited acoustic stress pattern with many intermediate stressed syllables and not many primary stresses. This flattens the scatter plot, reporting many informative words mainly on the $y = 0.5$ line and confusing them with other kinds of words with intermediate stress. Moreover if the content of the monologue is not clear the grouping of the words in the two informative-stressed and uninformative-unstressed groups may result quite confused. In this case, in fact, the keywords are not easy to define, because the information to transfer to the listener is missing.

Based on the previous considerations, it is not easy to define a general frame that describes the relationship between information amount and average stress of each word inside each monologue. Some more global analysis than the scatter plot is needed. In this section, the histograms are reported for different kinds of words towards either their IR index (tf_idf) or their average acoustic stress. It is possible in this way to infer the distribution of a particular category of words, for example words with high average acoustic stress or words with high tf_idf, with respect to their IR index the first ones and to their average acoustic stress the second ones.

In figures 8.a and 8.b the proportion of words with average acoustic stress higher than 0.6 is considered for different tf_idf intervals for both transcribers' data respectively. The threshold 0.6 was chosen after the observation of the scatter plots in the previous section. In fact, words tend to be located on the three main average stress lines: $y = 0.0$, $y = 0.5$ and $y = 1.0$. Assuming that the keywords are the most stressed words, a threshold that excludes unstressed and only mildly stressed words must be used.

In both figures, more in figure 8.a than in figure 8.b however, the ascending trend of the proportion of words is clearly detectable. A very low proportion of words with average acoustic stress higher than 0.6 has a tf_idf value close to 0.0 or more in general very low. On the opposite the higher proportion of words with average acoustic stress above 0.6 becomes higher and higher with the increasing of the tf_idf value. For the highest tf_idf values (tf_idf between 13.0 and 14.0), the proportion of words with average acoustic stress > 0.6 becomes 1.0 for the data labelled by transcriber # 1 (figure 8.a) and saturates to 0.7 for transcriber # 2's data (figure 8.b).

These two histograms show that a word with high information content, as determined by the tf_idf index, has a high probability of having an average stress above 0.6. It has to be considered that the distribution of words for high values of tf_idf shows pretty low values (Fig. 7.a and 7.b) and consequently the 1.0 proportions on the right part of the histograms are based only on a small number of instances. The ascending trend, however, is observed also below these extreme values of tf_idf.

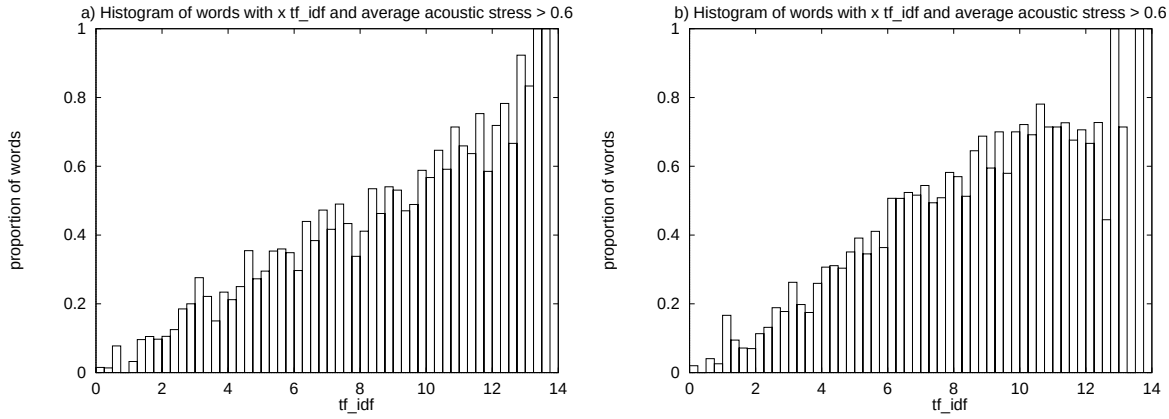

Figure 8: Histogram of words with average acoustic stress > 0.6 vs. tf_idf a) for the first and b) for the second transcriber's data.

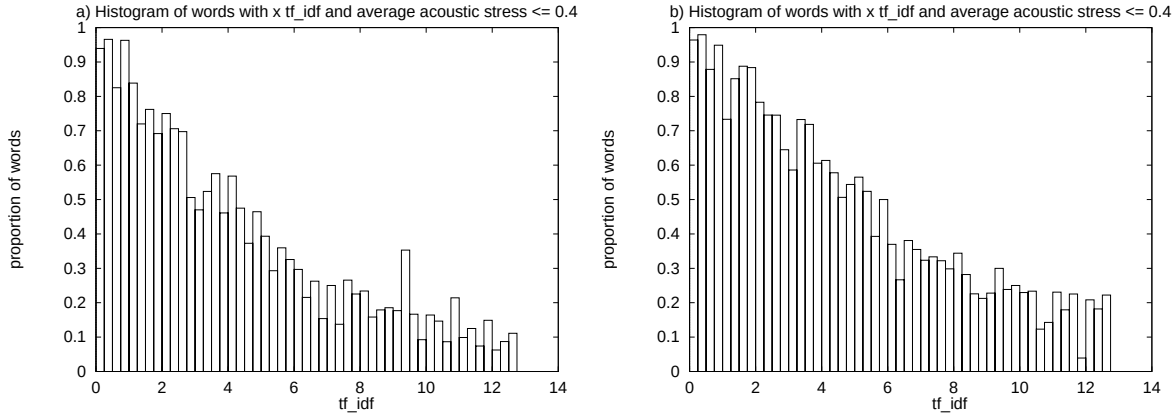

Figure 9: Histogram of words with average acoustic stress $<= 0.4$ vs. tf_idf a) for the first and b) for the second transcriber's data.

The complementary histograms to the ones shown in figure 8.a and 8.b report the proportion of words with average acoustic stress lower than 0.4 across all the two transcribers' data sets and are depicted in figure 9.a and 9.b. As it was to be expected the two histograms show a complementary trend to the ones in figures 8.a and 8.b. Words with low tf_idf values, that is with low information gain, are likely to have an average acoustic stress lower than 0.4 in each monologue (almost 1.0 proportion of words in both histograms). On the opposite words with high information content, for example with tf_idf above 8.0, are unlikely to have a low average acoustic stress.

Between the 0.6 and 0.4 thresholds fixed on the average acoustic stress in figures 8.a and 8.b and 9.a and 9.b there are a number of words with neither low nor high average acoustic stress value. It would be interesting to locate such words in terms of conveyed information. The two histograms for this middle class of words are reported in figure 10.a and 10.b for transcriber # 1's and transcriber # 2's data respectively. All words in this case concentrate around middle values of the IR index. This shows that words with medium values of average acoustic stress are also characterized by medium information content.

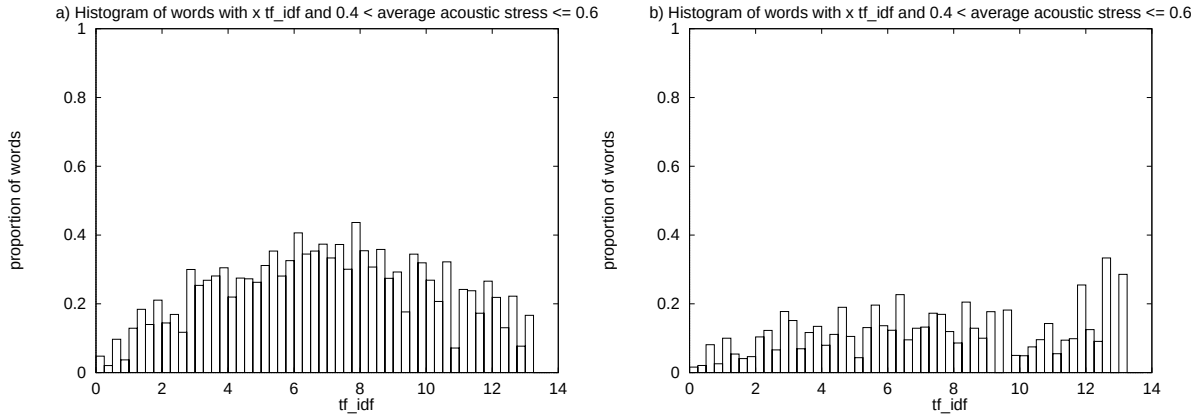

Figure 10: Histogram of words with 0.4 < average acoustic stress <= 0.6 vs. tf_idf a) for the first and b) for the second transcriber's data.

In the previous figures we have seen that a word with high information content is also likely to have a high average value and on the opposite that a word with low information content is very likely to be unstressed across the whole monologue. Is it possible to state the opposite, that is that a word with high average acoustic stress is very likely to be a keyword? To answer this question, the histogram of the words with tf_idf higher than 6.0 is drawn towards the average acoustic stress values (figures 11.a and 11.b). In this case the threshold 6.0 was chosen quite arbitrarily to distinguish words with high information from words with low information.

The distribution of words across the average acoustic stress values was not very homogeneous, being mainly concentrate at 0.0, 0.5 and 1.0 (Fig. 6.a and 6.b). The number of words with a different value of average acoustic stress is not sufficient for a significant estimation of the corresponding word probability. More reliable conclusions might be derived looking at the histograms bars located at the only average acoustic stress values with a statistically significant number of words: 0.0, 0.5 and 1.0. After isolating these three bars from the two histograms, an ascending trend may be observed. The proportion of words with tf_idf higher than 6.0 with average acoustic stress 0.0 is lower than the same proportion of words with average acoustic stress 0.5, which is lower than the same proportion with average acoustic stress 1.0. This ascending trend seems to confirm that a stressed word is likely to be also informative. However the limited number of statistically significant values of average acoustic stress does not allow sure conclusions.

5 Conclusions

The relationship between acoustic stress and IR weighting indexes has been investigated. For each spoken document, words have been characterized on one side by means of their average acoustic stress throughout the monologue and on the other side by using a heuristic measure of information widely accepted in the IR literature.

First the scatter plots of the average acoustic stress vs. the IR index were drawn. The majority of the analyzed files produced a scatter plot where two groups of words were clearly distinguishable: on the bottom left corner words with no informative value and low average acoustic stress and on the top right corner words with high IR index and high average stress are grouped together. Between

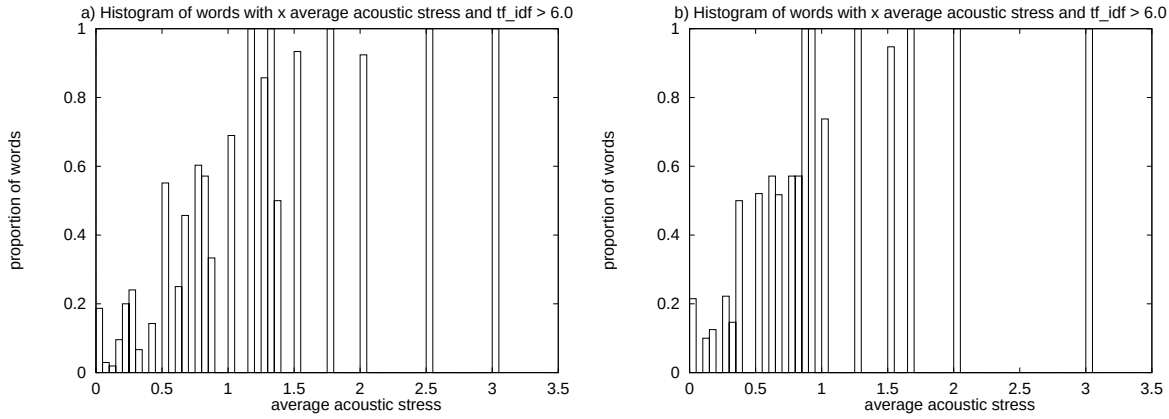

Figure 11: Histogram of words with $tf\_idf > 6.0$ vs. average acoustic stress a) for the first and b) for the second transcriber's data.

these two groups of words there is a third group of mildly informative and mildly stressed words. In some cases the words in the scatter plot do not assume this configuration. This can happen because the speaker does not talk about a defined topic, and therefore it is hard to define keywords, or because the speaker does not make much use of prosodic stress, but speaks in a very monotonous way.

The scatter plots show that for many speakers there is some correspondence between acoustically stressed and informative words, but does not show a global picture of such correspondence. Thus the histogram of highly stressed words vs. the IR index is calculated. Even here we can see that words with high average stress are mainly concentrated at the high values of the IR index. The inverse histogram of highly informative words vs. their average acoustic stress can be considered statistically significant only for three levels of average stress and consequently no sure conclusions can be derived.

This study confirms the hypothesis of a direct relationship between acoustic stress and information in spoken sentences, based on a large amount of data. It also produces some statistical figures to characterize such a relationship.

Next step will be to integrate acoustic stress and IR indexes into a topic detection weighting algorithm that will take into account both acoustic and statistical clues. This new algorithm will be extremely useful in a number of tasks and in particular in topic detection and tracking in spoken documents. This will be the subject of future work.

References

[1] K. Silverman, M. Beckman *et al.*, "ToBI: A Standard for Labeling English Prosody", *International Conf. on Spoken Language Processing*, 1992.

[2] I. Lehiste, *Suprasegmentals* MIT Press, Cambridge, 1970.

[3] J. Clark and C. Yallup, "Introduction to Phonology and Phonetics", Oxford: Blackwell, 1990.

[4] M. Horne, P. Hansson, G. Bruce, J. Frid, "Prosodic Correlates of Information Structure in Swedish Human-Human Dialogues", *Proc. Eurospeech*, pp. 29-32, 1999.

[5] S.E. Robertson and K. Spark Jones, "Simple, proven approaches to text retrieval", Technical Report TR356, Computer Laboratory, University of Cambridge, UK, May 1997.

[6] K. Spark Jones, S. Walker, and S.E. Robertson, *A probabilistic model of information retrieval: development and status*, Technical Report TR446, Computer Laboratory, University of Cambridge, UK, August 1998.

[7] D. Kuijk, L. van and Boves, 1999. Acoustic characteristics of lexical stress in continuous telephone speech. *Speech Communication* 27, 95-111.

[8] C.W. Wightman, and M. Ostendorf, 1994. Automatic labeling of prosodic patterns. *IEEE Transactions on Speech and Audio Processing* 2,469-81.

[9] R. Silipo and S. Greenberg, "Automatic transcription of prosodic stress for spontaneous English discourse", Proc. of the XIVth International Congress of Phonetic Sciences (ICPhS), 3:2351, 1999.

[10] D. van Bergem, 1993. Acoustic vowel reduction as a function of sentence accent, word stress and word class on the quality of vowels. *Speech Communication* 12, 1-23.

[11] Center for Spoken Language Understanding, Dept. of Computer Science and Engineering, Oregon Graduate Institute. 1995. *Stories corpus*, Release 1.0.

[12] N. Chomsky and M. Halle, "The Sound Pattern of English." New York: Harper and Row, 1968.

[13] Manning, Schutze. "Foundations of statistical natural language processing". MIT Press. 1999.

[14] F. Crestani, "Vocal access to a newspaper archive: design issues and preliminary investigation", in: *Proceedings of ACM Digital Libraries*, pp. 59–68, Berkeley, CA, USA, August 1999.

[15] W.B. Frakes, "Stemming algorithms", in: W.B. Frakes and R. Baeza-Yates, editors, *Information Retrieval: data structures and algorithms*, chapter 8. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1992.

[16] H.P. Luhn, "The automatic creation of literature abstracts", *IBM Journal*, pp. 159–165, April 1958.

[17] M.F. Porter, "An algorithm for suffix stripping", *Program*, 14(3):130–137, 1980.

[18] G. Salton, *Automatic information organization and retrieval*, McGraw Hill, New York, 1968.

[19] A. Tombros and F. Crestani, *Users's perception of relevance of spoken documents*, Technical Report TR-99-013, International Computer Science Institute Berkeley, CA, USA, July 1999.

[20] A. Tombros and M. Sanderson, "Advantages of query biased summaries in Information Retrieval", in:*Proceedings of ACM SIGIR*, pp. 2–10, Melbourne, Australia, August 1998.

[21] C.J. van Rijsbergen, *Information Retrieval*, Butterworths, London, second edition, 1979.

A Scatter Plots of Speech Files Labelled by Transcriber # 1

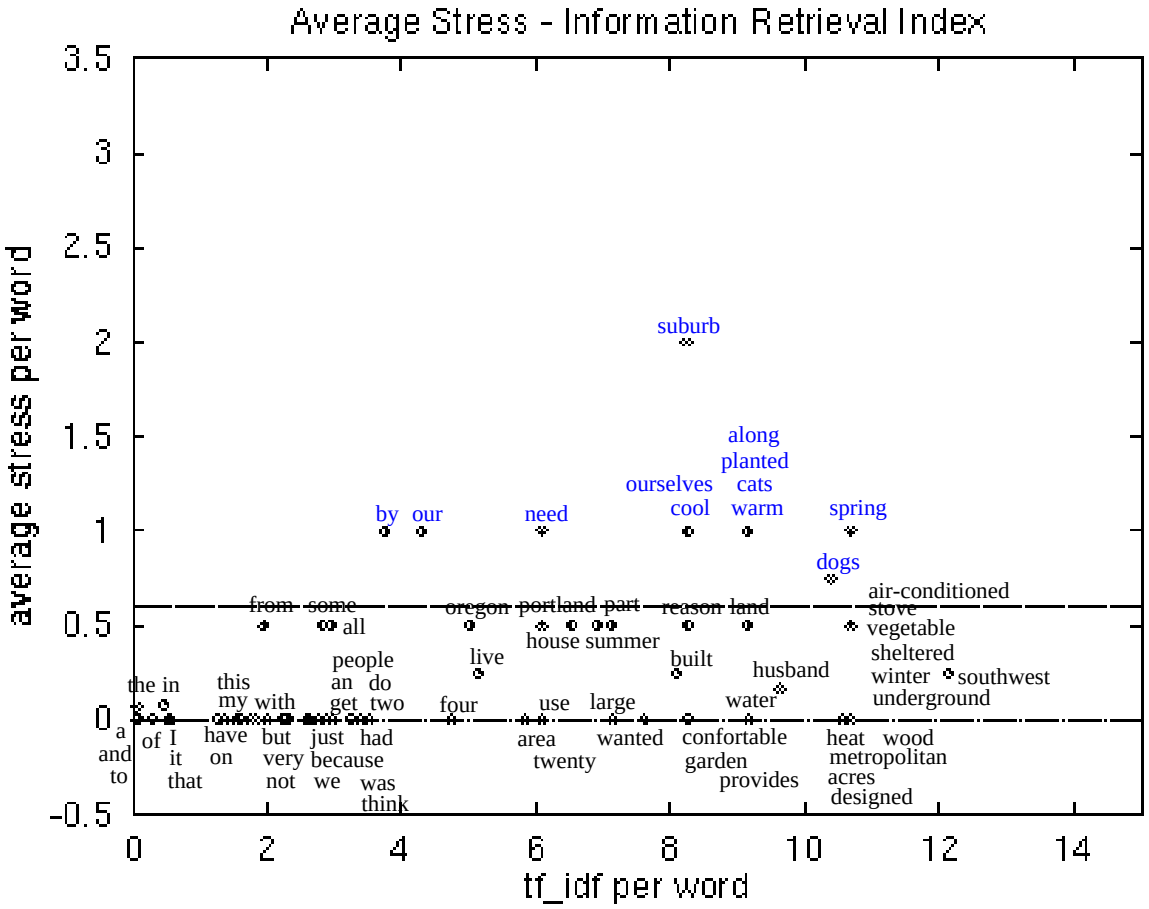

Figure 12: ".pau i live(+2) in a suburb(+1+1) .pau in the southwest part(+1) .pau of .pau oregon(+2) .pau or southwest(+2) part of the metropolitan area of portland(+2) .pau .br .pau with my husband(+2) .pau .br .pau four dogs(+2) .pau two dogs(+1) .pau .ls and .pau four cats(+1) .pau .ls .pau we live in a house(+2) that my husband designed .pau and we built ourselves(+1) .pau it is an underground(+2) .pau or earth sheltered(+2) house(+2) .pau .br .pau the(+2) reason(+2) we built(+2) it was because .pau my husband had wanted to do this all(+2) along(+1) .pau .br .ls .pau we live on twenty acres of land(+2) .pau we have a spring(+1) that we get our water from(+2) .pau .br .pau and we have a large vegetable(+2) garden that we of just planted(+1) .br .pau .ls .br the house(+2) is very comfortable to live(+2) in(+2) .pau .br .pau it .pau is .pau warm(+1) in the winter(+2) and cool(+1) in the summer(+2) .pau some(+2) people think it's airconditioned(+1) but it's not .pau .br .ls .pau we use a wood stove(+2) to heat by(+1) .pau .ls .pau and it .pau heats provides all(+2) the heat .pau we need(+1)". (+1) = primary stress; (+2) = intermediate stress.

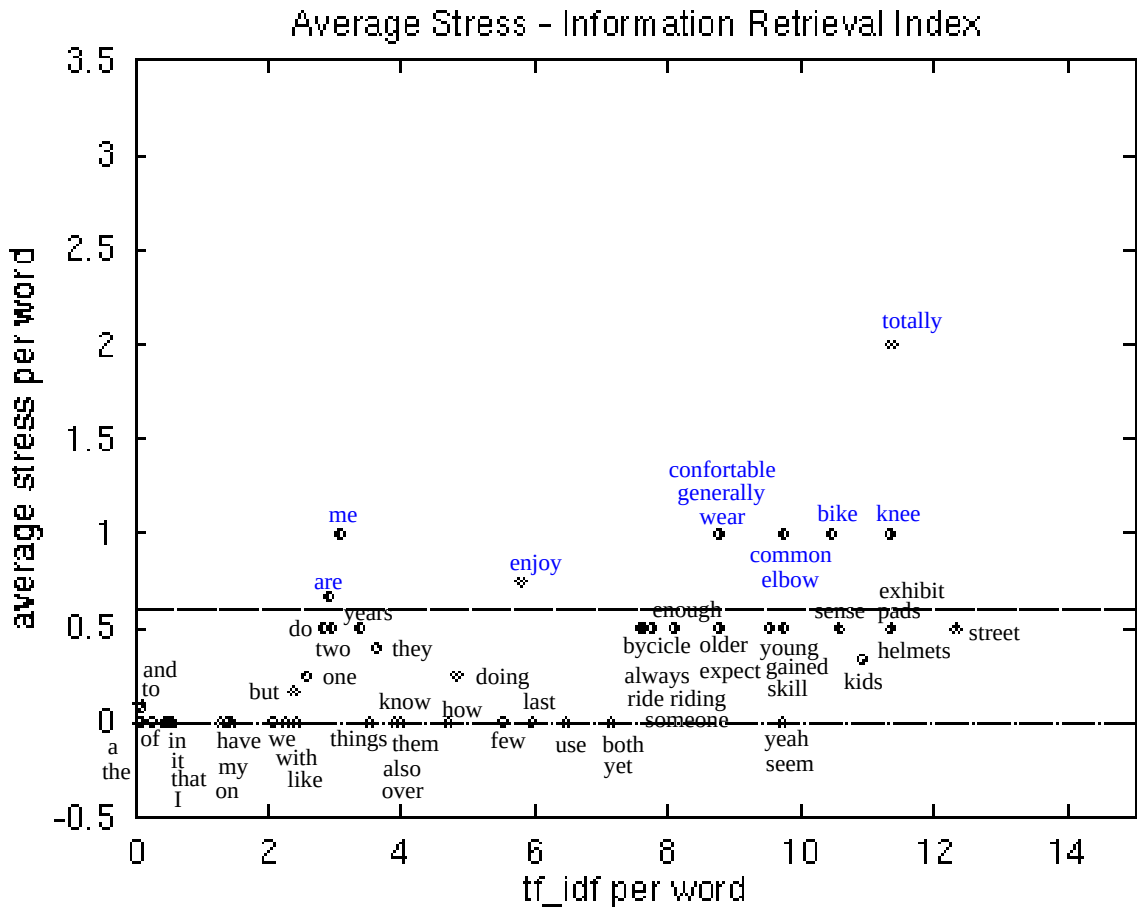

Figure 13: ".pau yeah one(+2) of things i like to do(+2) is to .pau ride(+1) with bikes(+1) with my kids(+1) .pau my kids are young(+2) enough(+2) .pau uh .uu the youngest(+2) one doesn't ride(+2) a bike yet he's to young(+2) .pau .br .pau but(+2) .pau the older(+2) two(+2) are(+1) uh .pau they both know how to ride a bike(+2) they .pau have .pau uh .pau .br .pau .ls .pau .ls .pau gained(+2) that skill(+2) over the last few years(+2) .pau .ls .br .pau uhm .pau we always(+2) use bicycle(+2) helmets(+2) and .pau and uh .pau i generally(+1) we're riding(+2) on street(+2) i have .pau my kids also wear(+1) .pau .br elbow(+2+2) .pau and(+2) .pau knee(+1) pads(+1) .pau .br .pau .br .pau uhm .pau .ls .br .pau .ls they(+1) are(+1) .pau totally(+1+1) .pau uh .br .laugh comfortable(+1) with riding(+2) in the street(+2) and in .pau in a sense that .pau .br .pau they(+1) don't(+1) .pau exhibit(+2) a lot(+1) a common(+1) sense(+1) that i .br expect(+2) someone(+2) to(+2) but .pau .br .pau but uh .pau i enjoy(+1) doing(+2) it with .pau them and they seem to enjoy(+2) doing it with .pau me(+1) .pau ". (+1) = primary stress; (+2) = intermediate stress.

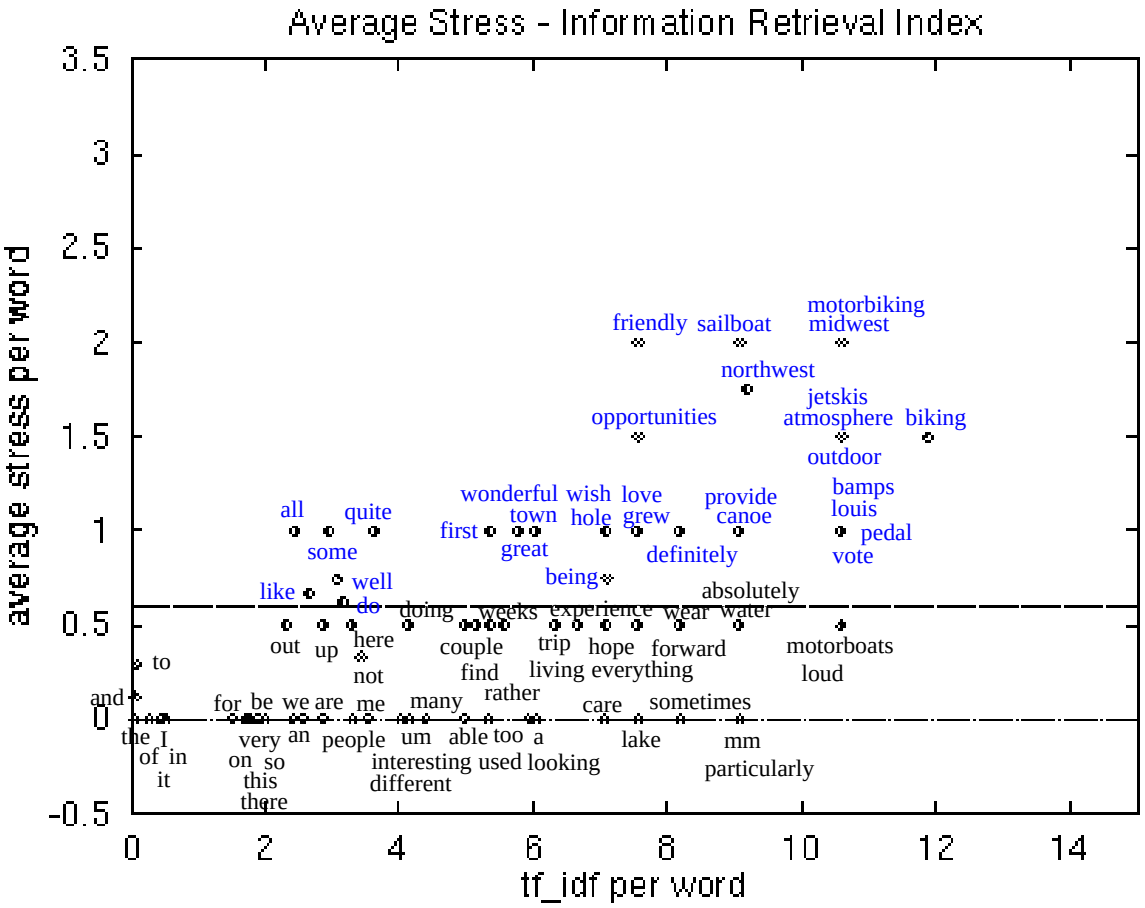

Figure 14: ".pau well(+2) this is this is a rather(+2) interesting assignment(+1) i'm not(+2) uh i'm not used to doing(+2) this .pau .br .pau well(+1) first(+1) of all(+1) i'm living(+2) in an absolutely(+2) wonderful(+1) town(+1) in the northwest(+1+1) .pau .ls .br .pau it provides(+1) a great(+1) .pau many opportunities(+2+2+2) i do(+1) wish(+1) sometimes .pau we didn't vote(+1) on everything(+2) quite(+1) as .pau much as .pau we do(+2) .pau .br .pau um i like(+1) to(+1) .pau .br .pau do(+2) things on the water(+2) .pau mm like canoe(+1) and sailboat(+1+1) .pau don't(+1) particularly care for motorboats(+2) there a little too loud(+2) and jetskis(+1+2) are definitely(+1) out(+1) .pau .br .pau i'm(+2) looking forward(+2) to(+1) a trip(+2) to bamps(+1) and lake louis(+1) in a couple(+2) of weeks(+2) .pau .br .pau hope(+2) to do(+2) some(+1) biking(+1+1) .pau pedal(+1) biking(+1) .pau not(+2) .pau motorbiking(+1+1) .pau .ls .br .pau and(+2) i grew(+1) up(+2) .pau in the midwest(+1+1) so being(+1) out here(+2) in the northwest(+1+2) is a whole(+1) different experience(+2) for me .pau .br .pau i find(+2) the people to be very friendly(+1+1) .pau .br .pau i love(+1) the outdoor(+1+2) atmosphere(+2+1) .pau .br .pau i like(+1) being(+2) able to wear(+2)". (+1) = primary stress; (+2) = intermediate stress.

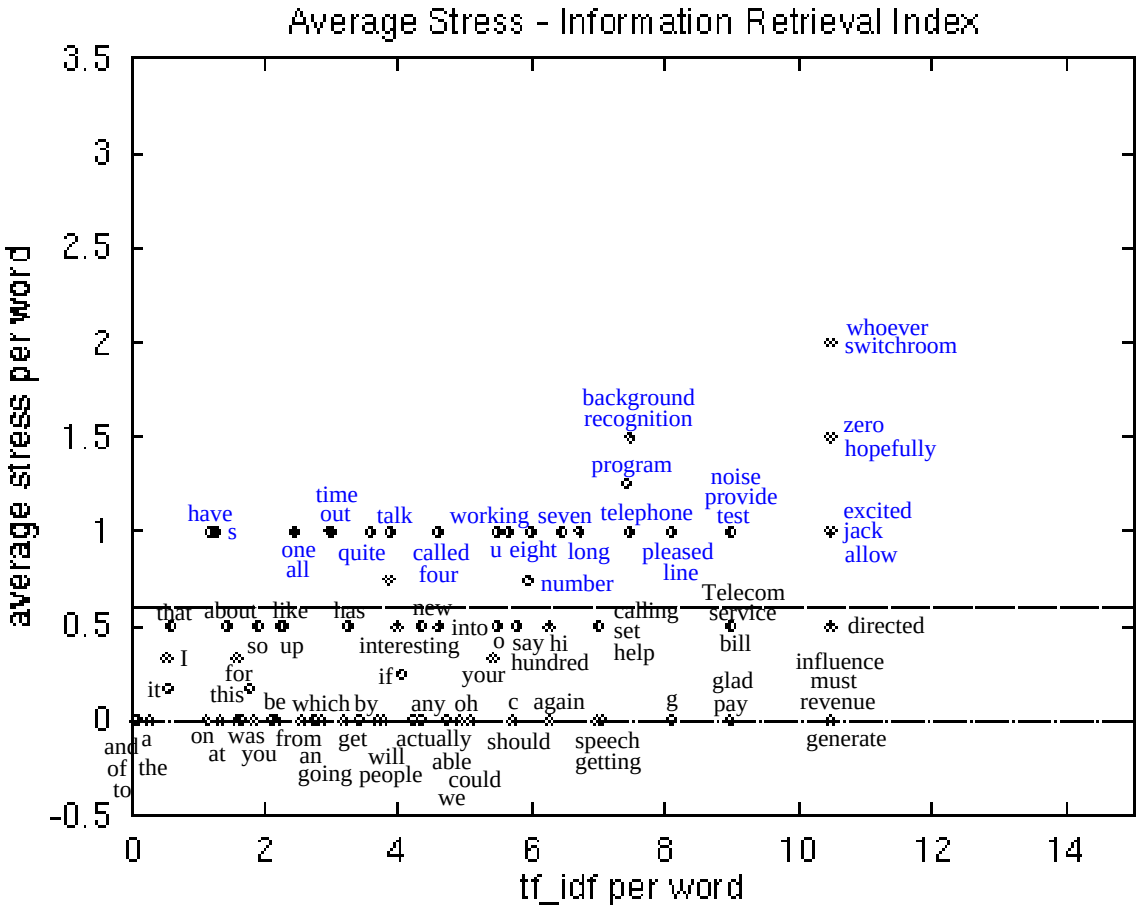

Figure 15: ".pau oh hi(+2) this is jack(+1) from o(+2) g c telecom(+2) actually i called(+1) out(+1) to test(+1) .pau out(+1) this new(+2) eight(+1) hundred(+2) number(+2+2) .pau .br to see(+2) if it was working(+1) which it is(+1) it's eight(+1) hundred(+2) four(+1) four(+1) one(+1) one(+1) zero(+1+2) seven(+1) seven(+1) .br .pau i must say(+2) i'm quite(+1) pleased(+1) as we uh provide(+1) your telephone(+1) service(+2) to be able to have(+1) an eight(+1) hundred(+2) number(+2) directed(+2) into(+2) a .pau program(+1+2) like(+2) this(+2) which uh is going to allow(+1) people to talk(+1) for a long(+1) long(+1) time(+1) .pau that'll generate lots(+2) of revenue for us(+1) and .br .pau hopefully(+2+1) uh .pau whoever(+1+1) .pau uh u s(+1) west(+1) will get to pay the bill(+2) for(+1) it(+2) .pau .br so(+1) were excited(+1) about(+2) that(+1) .pau again(2) i'm calling(+2) from the switchroom(+1+2) so there's a lot of background(+1+2) noise(+1) which should be interesting(+2) to see(+1) .pau .br .pau if(+2) that(+2) has(+2) any influence at all(+1) on your(+1) uh your speech recognition(+2+1) .pau program(+1) .pau .br .pau uhm .pau i'm glad we could .pau .br .pau help(+2) .pau you out(+1) by getting that(+2) uh .br .pau getting that line(+1) set(+2) up(+2) .pau". (+1) = primary stress; (+2) = intermediate stress.

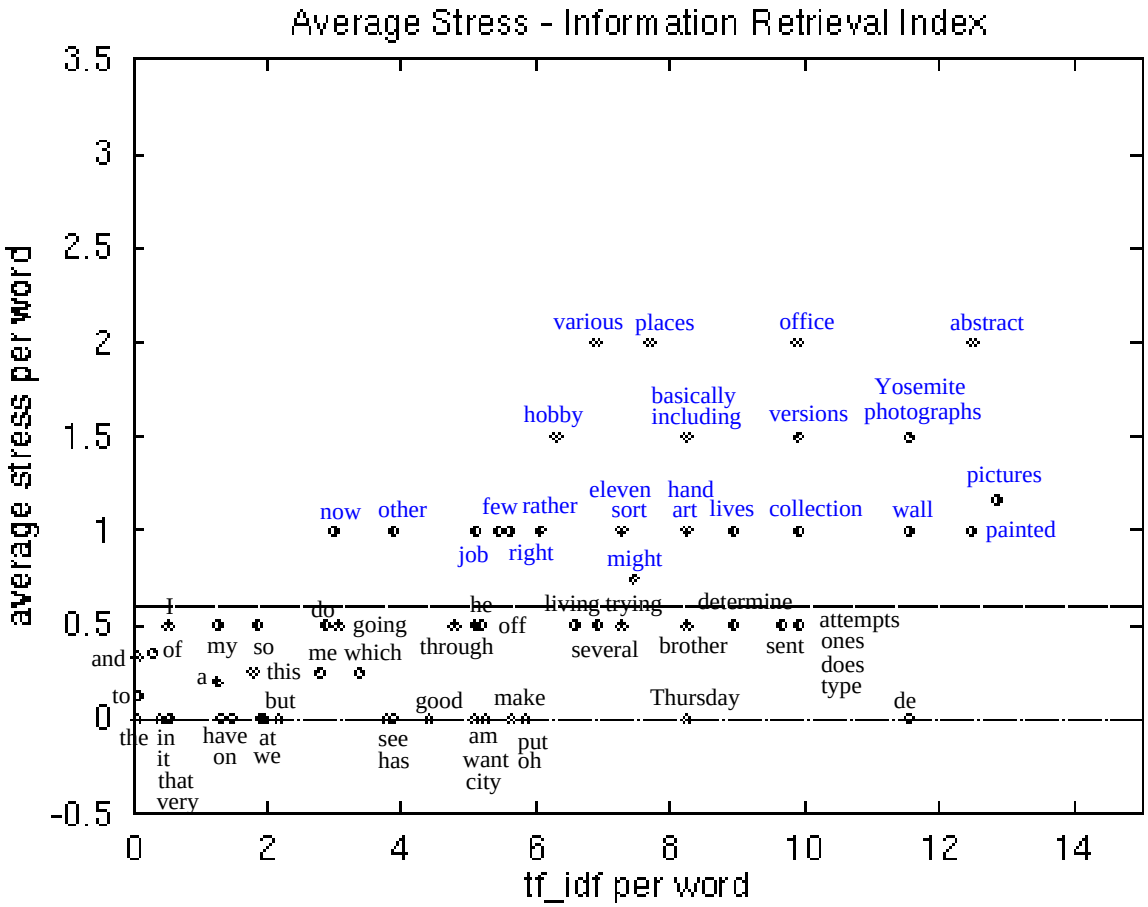

Figure 16: "Right now I'm in my office it's eleven o'clock at night .br thursday night I am going through a collection of pictures that my brother has sent me which he has painted .br trying to de determine which one I want to put on the wall .ls uhm .br see we have various abstract pictures including a few .br uh hand painted versions of photographs of Yosemite and several .br other places uhm city of type of art is .br rather abstract and uh .br he basically .br lives at home and does this .br as sort of a uh hobby but attempts to make a living off it doesn't do a very good job so he sent me to these .br pictures oh of".

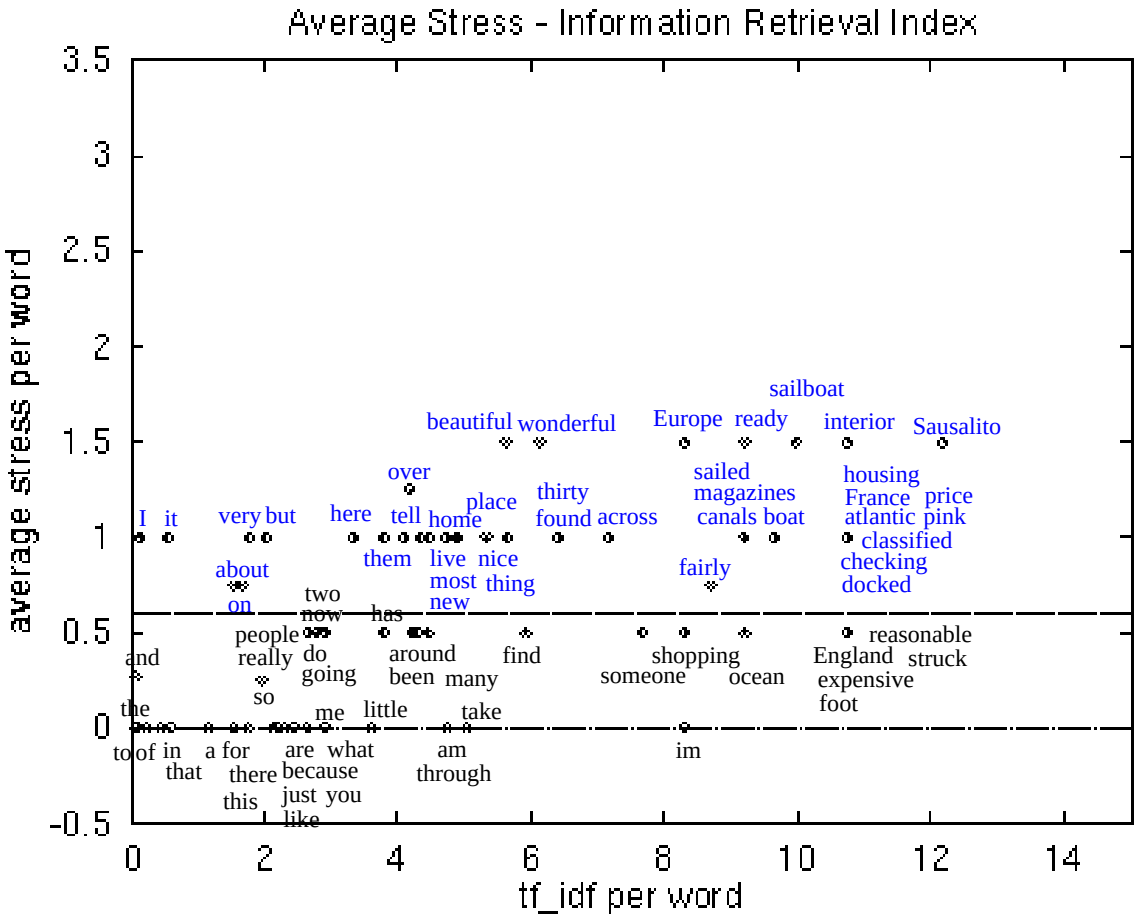

Figure 17: "I'm going to tell you about shopping for a sailboat .br uhm because I am uh fairly new to Sausalito and the thing .br that struck me the most .br wonderful about this place that there are so many nice sailboats around .br lots of people live on them .br and that what I'd like to do here because it's very expensive to uhm find housing .br but you can find a nice boat and a fairly reasonable price and live on it .br I have been uh checking the the classified boat magazines and I have found a really nice .br thirty two foot sailboat .br that has a beautiful pink interior .br and uhm has been sailed across the Atlantic ocean over to Europe and through canals in France and England .br and now it's docked in Sausalito and just ready .br for someone to take over as a little home".

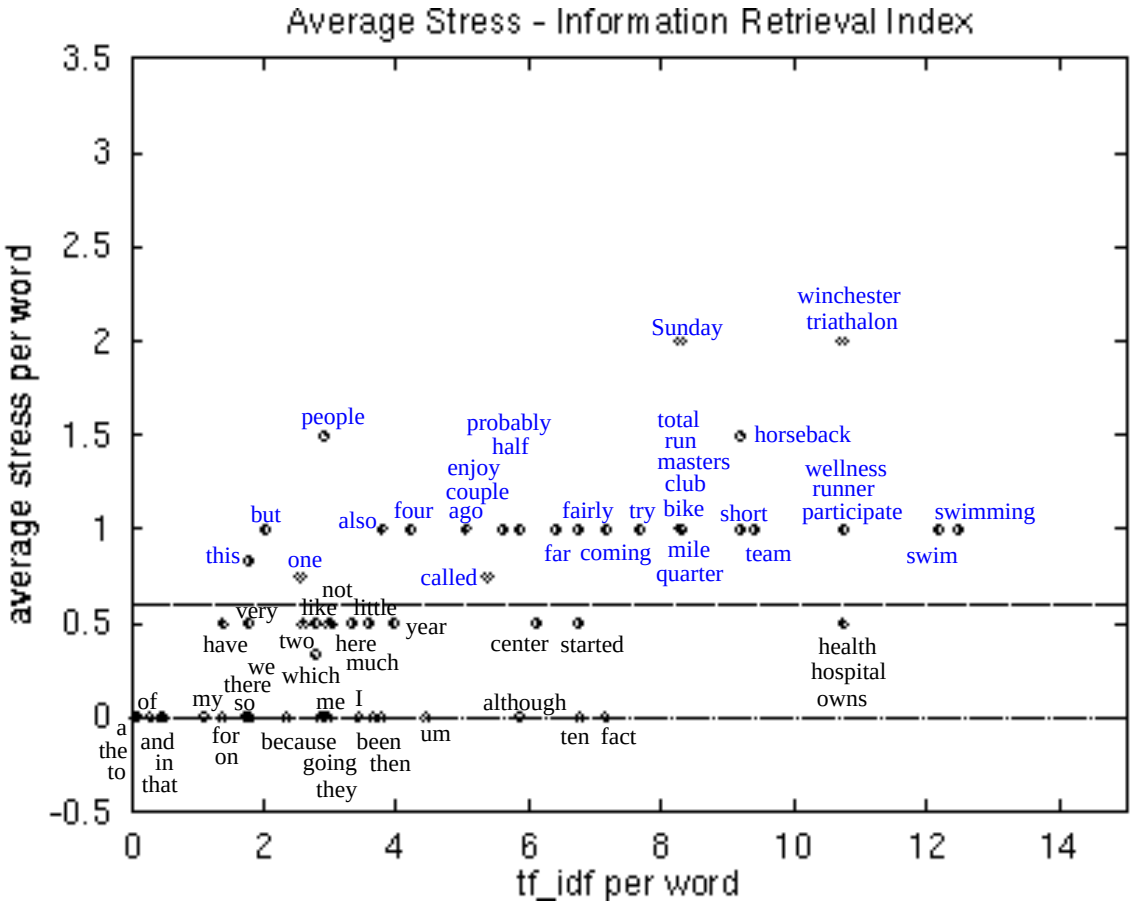

Figure 18: "I really enjoy swimming I started a couple years ago we have a what's called a wellness center here which is really a health club that the hospital owns .br and I've been swimming there fairly regularly they have what's called a master swim team .br um and this a lot of people on my swim team like to participate in triathalon and so I'm going to try one .br this coming sunday in Winchester it's a rather short one but I'll swim a quarter of a mile which isn't very far and then bike ten miles .br and .ls run .br two and half miles .ls which is a little far for me because I'm not much of a runner in fact this year I have probably .br ran a total of four miles .ls .br I also like to uh .br horseback ride although I".

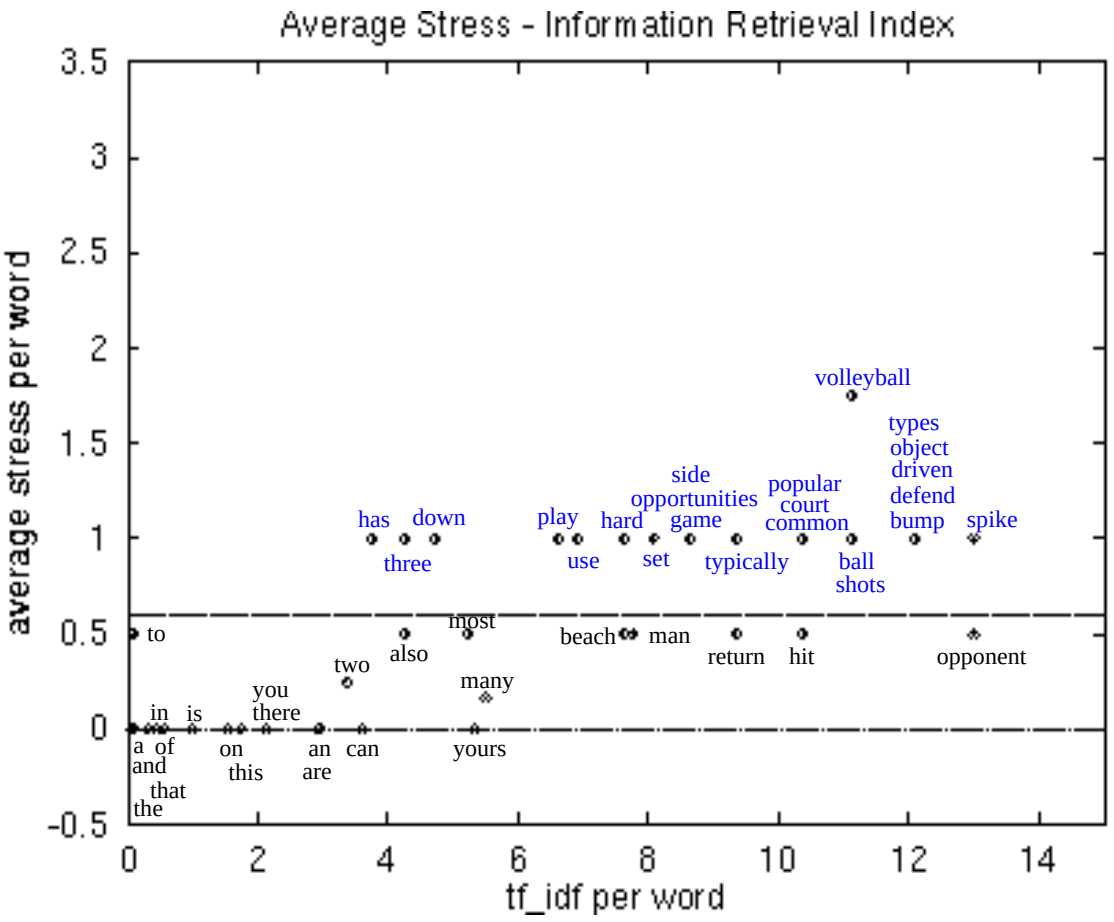

Figure 19: "Two men beach volleyball .br .ls .br uh .br the object of the game .br .br is to .br .br uh .br it's to .br uhm hit the ball down on your opponent's side of the court .br .br uh the opponent has .br .br three opportunities .br .br to defend .br and return the ball .br .br typically .br .br this is a bump .br a set and a spike .br uh .br there are many .br .br many types of shots .br .br that you can use in two man volleyball .br uhm .br most common and most popular is a hard driven spike .br there are .br .br also uhm .br many play shots".

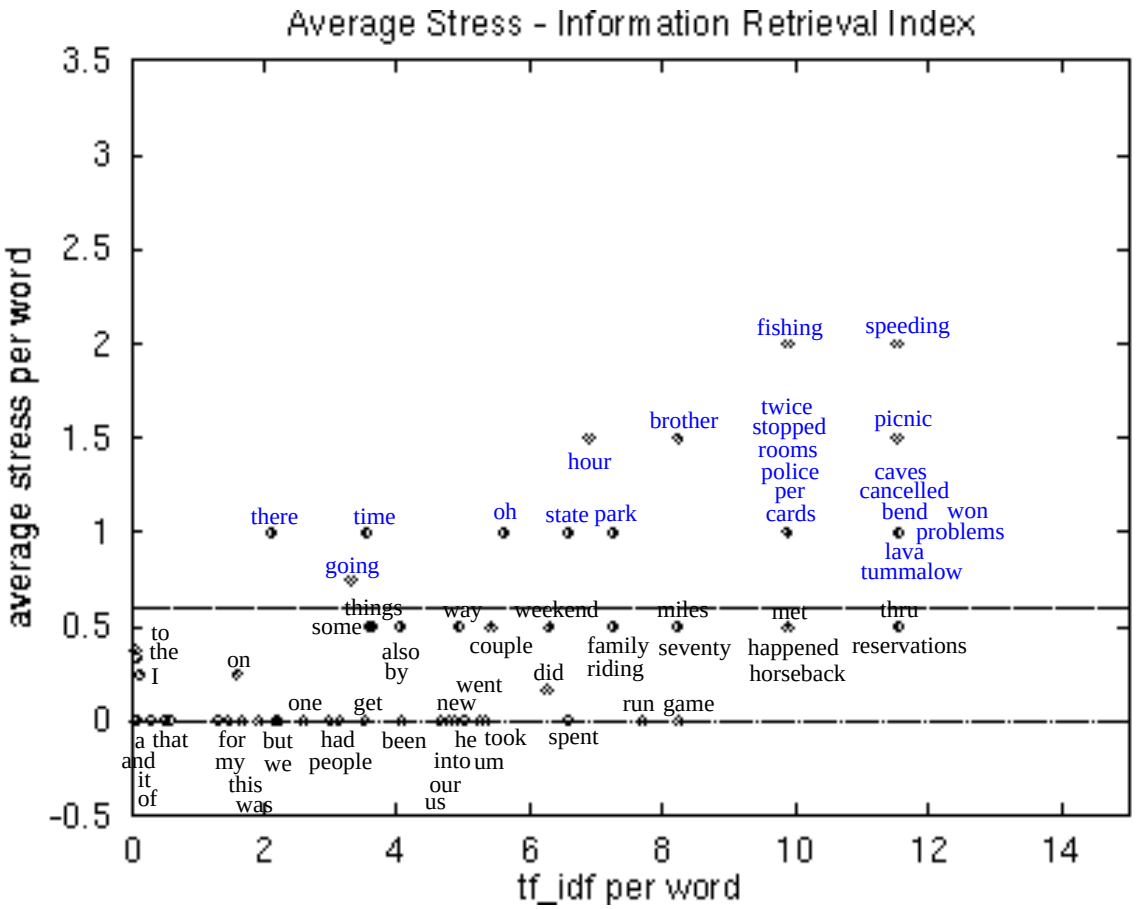

Figure 20: "This weekend I went to Bend and met my family there um we spent some time going thru the lava caves we also did some horseback riding some of the people there did some fishing we went on a picnic to Tummalow state park um oh one of the things that happened was that I won a game of cards um we did run into a couple of problems our reservations had been cancelled but it took us some time to get new rooms and on my way there my borther was stopped twice by the police for speeding he was going seventy one miles per hour and he ".

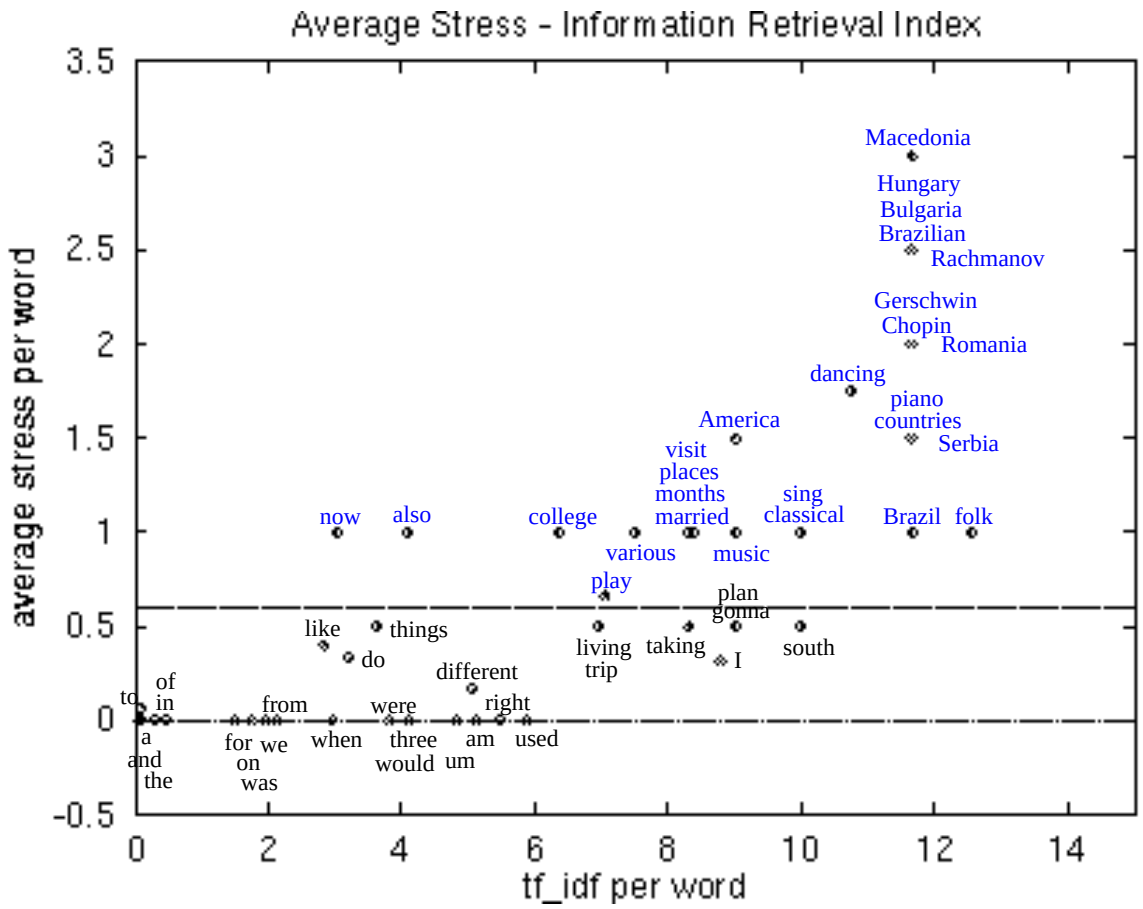

Figure 21: "I like to do a lot of different things when I was in college I used to do a lot of folk dancing from various different countries Bulgaria Macedonia Serbia Romania Hungary Israel uh I like to sing I also play the piano I like to play classical music I like to play a lot of Chopin Rakmaninonv Gershwin um I I like living in different places right now I am married to a Brazilian and we plan on taking a trip to Brazil for three months we're gonna visit various places in South America and".

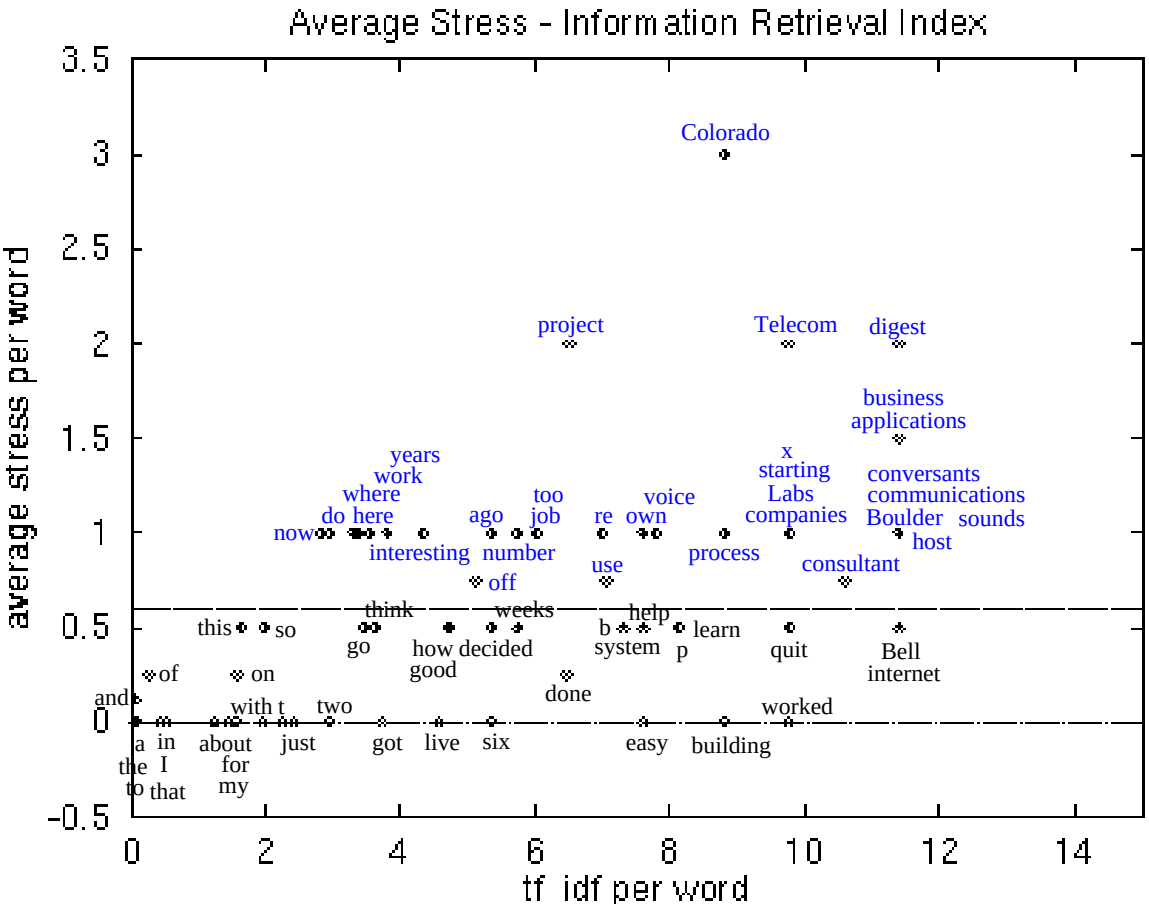

Figure 22: "I live in Boulder Colorado and I've worked here for about six years .br as a consultant to A T and T Bell Labs and uh two weeks ago .br decided to quit my job and go off on my own as a .br communications consultant .br just where I got your number .br of the Internet Telecom digest .br so I'm in the process now of starting a business to help companies .br do PBX to host applications .br and learn how to use voice response systems .ls I've done a lot of work with the conversants .br and I think this project is interesting .br you've done a good job on .br building the voice response system so that it's easy to use .br and sounds good too".

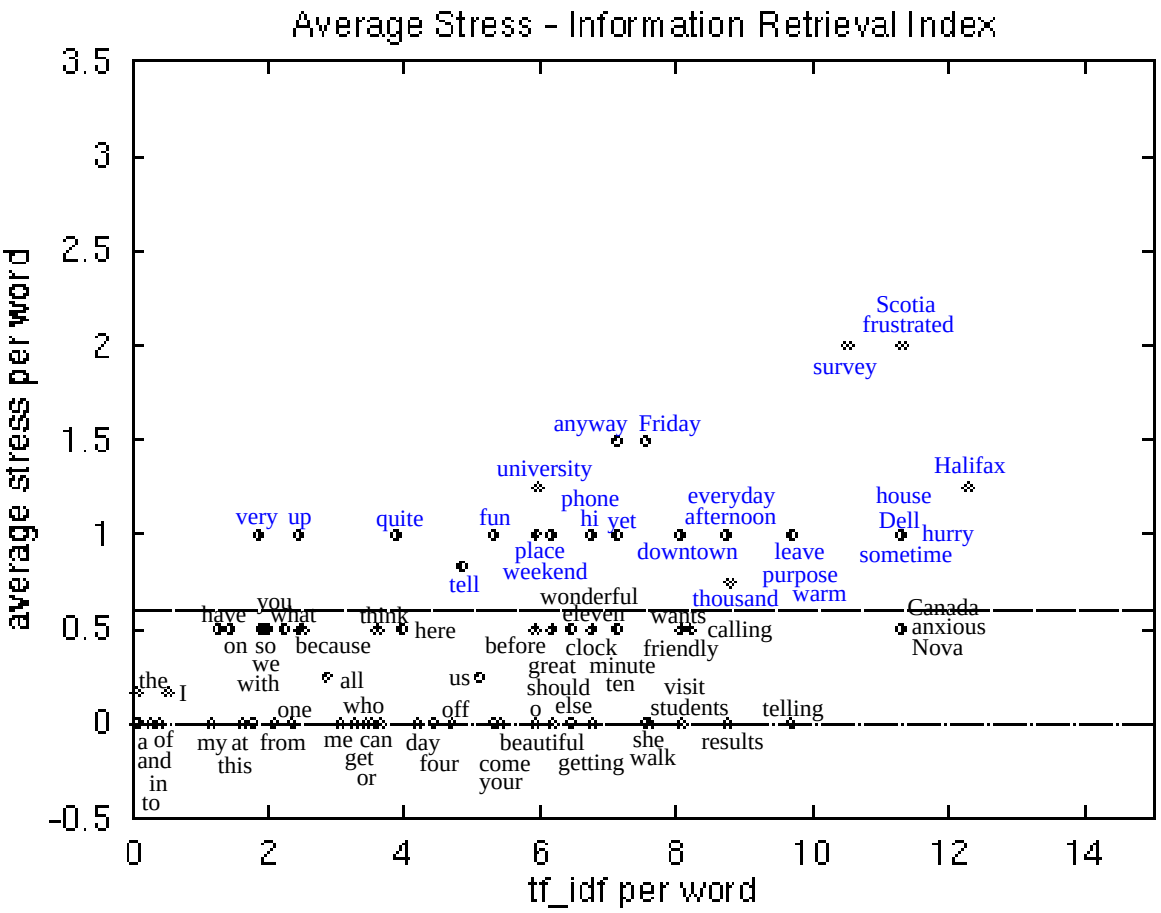

Figure 23: "Hi I'm calling it's four o'clock on a friday afternoon before the Canada day weekend here in Halifax Nova Scotia and one of my co-workers who I walk downtown everyday with is .br getting quite anxious because it's .br very warm and she wants to leave so she's frustrated she 's telling me to hurry up and get off the phone .br .br .br and uh .br what else can I tell you uh it's beautiful here at Halifax you should all come visit us some time tell us results of your survey .br calling from Dell house university .br a great university ten thousand eleven thousand students or so .br uhm wonderful friendly place we have here .br laugh .br this is fun .br what's the purpose of your survey anyway that's all I can think to tell you a minute yet".

B Scatter Plots of Speech Files Labelled by Transcriber # 2

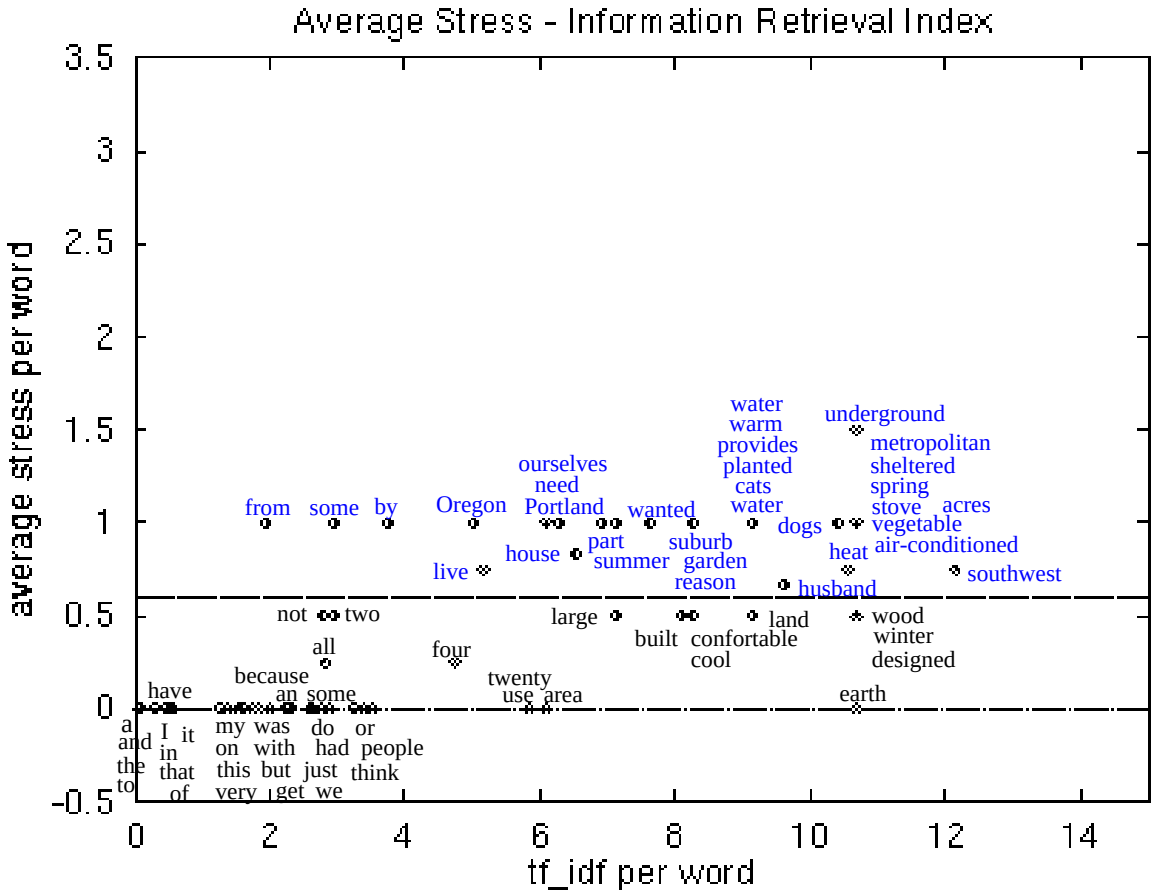

Figure 24: ".pau i live(+1) in a suburb(+1) .pau in the southwest(+2) part(+1) .pau of .pau oregon(+1) .pau or southwest(+1) part(+1) of the metropolitan(+1) area of portland(+1) .pau .br .pau with my husband(+1) .pau .br .pau four(+2) dogs(+1) .pau two(+2) dogs(+1) .pau .ls and .pau four cats(+1) .pau .ls .pau we live(+2) in a house(+1) that my husband(+2) designed(+2) .pau and we built ourselves(+1) .pau it is an underground(+2+1) .pau or earth sheltered(+1) house(+2) .pau .br .pau the reason(+1) we built(+1) it was because .pau my husband(+2) had wanted(+1) to do this all along(+1) .pau .br .ls .pau we live(+2) on twenty acres(+1) of land(+2) .pau we have a spring(+1) that we get our water(+1) from(+1) .pau .ls .br .pau and we have a large(+2) vegetable(+1) garden(+1) that we of just planted(+1) .br .pau .ls .br the house(+1) is very comfortable(+2) to live(+1) in .pau .br .pau it .pau is .pau warm(+1) in the winter(+2) and cool(+2) in the summer(+1) .pau some(+1) people think it's airconditioned(+1) but it's not(+2) .pau .br .ls .pau we use a wood(+2) stove(+1) to heat(+2) by(+1) .pau .ls .pau and it .pau heats provides(+2) all(+2) the heat(+1) .pau we need(+1)". (+1) = primary stress; (+2) = intermediate stress.

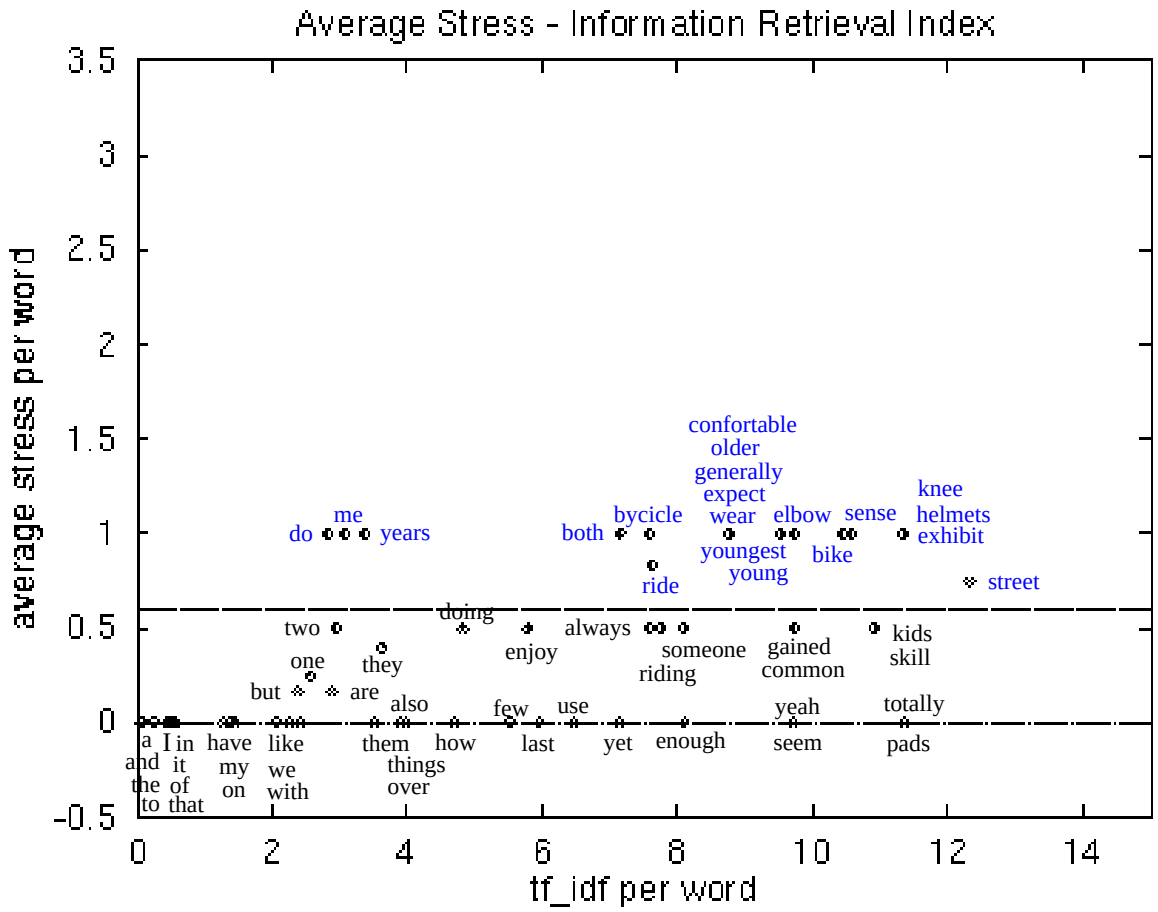

Figure 25: ".pau yeah one(+2) of things i like to do(+1) is to .pau ride(+2) with bikes(+1) with my kids(+1) .pau my kids(+2) are young(+1) enough .pau uh .uu the youngest(+1) one doesn't ride(+1) a bike(+2) yet he's to young(+1) .pau .br .pau but .pau the older(+1) two(+2) are uh .pau they both(+1) know how to ride(+1) a bike(+2) they .pau have .pau uh .pau .br .pau .ls .pau .ls .pau gained(+2) that skill(+2) over the last few years(+1) .pau .ls .br .pau uhm .pau we always(+2) use bicycle(+1) helmets(+1) and .pau and uh .pau i generally(+1) we're riding(+2) on street(+1) i have .pau my kids also wear(+1) .pau .br elbow(+1) .pau and .pau knee(+1) pads .pau .br .pau .br .pau uhm .pau .ls .br .pau .ls they(+1) are(+2) .pau totally .pau uh .br .laugh comfortable(+1) with riding(+2) in the street(+2) and in .pau in a sense(+1) that .pau .br .pau they(+1) don't(+2) .pau exhibit(+1) a lot a common(+2) sense(+1) that i .br expect(+1) someone(+2) to but(+2) .pau .br .pau but uh .pau i enjoy(+2) doing(+1) it with .pau them and they seem to enjoy(+2) doing it with .pau me(+1) .pau". (+1) = primary stress; (+2) = intermediate stress.

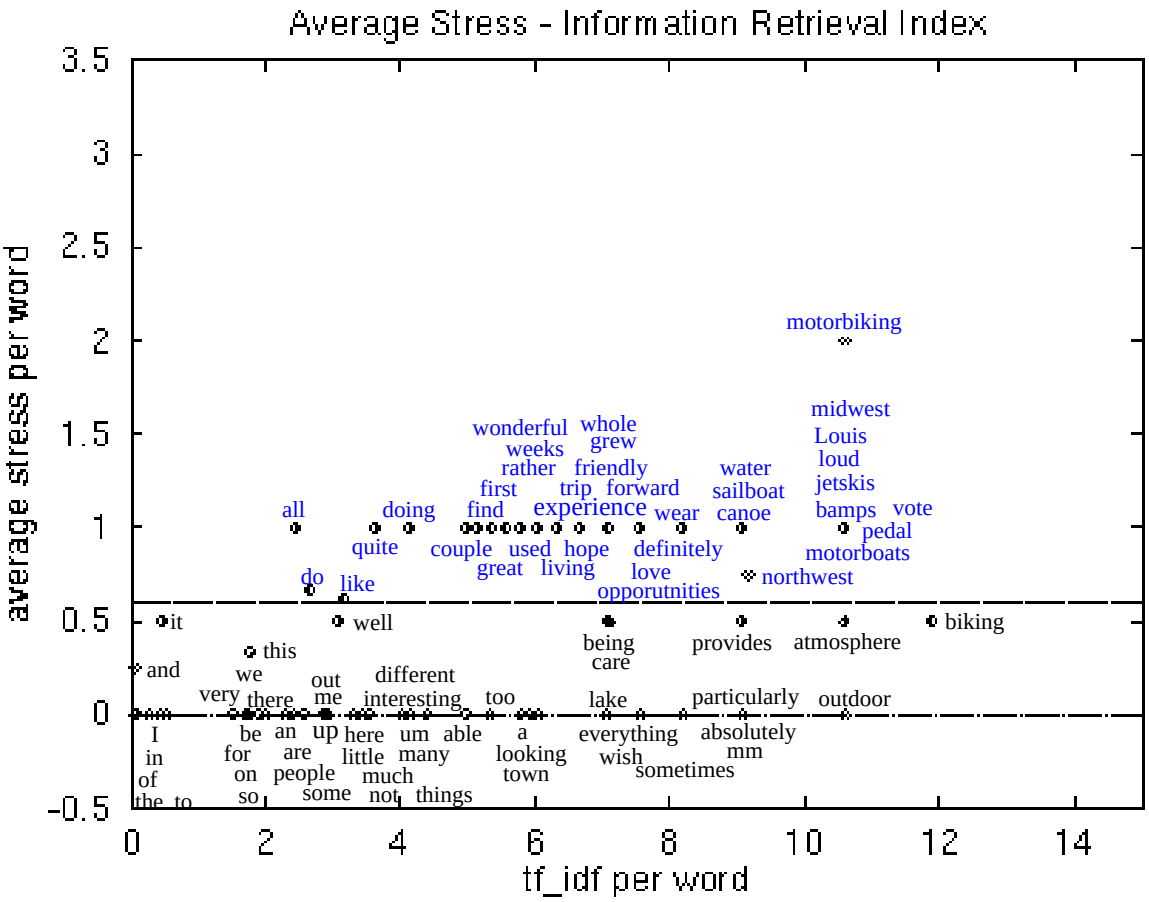

Figure 26: ".pau well this is this(+1) is a rather(+1) interesting assignment(+1) i'm not uh i'm not used(+1) to doing(+1) this .pau .br .pau well(+1) first(+1) of all(+1) i'm living(+1) in an absolutely wonderful(+1) town in the northwest(+2) .pau .ls .br .pau it(+2) provides(+2) a great(+1) .pau many opportunities(+1) i do(+1) wish sometimes .pau we didn't vote(+1) on everything quite(+1) as .pau much as .pau we do(+2) .pau .br .pau um i like(+1) to .pau .br .pau do(+1) things on the water(+1) .pau mm like canoe(+1) and sailboat(+1) .pau don't(+1) particularly care(+2) for motorboats(+1) there a little too loud(+1) and jetskis(+1) are definitely(+1) out .pau .br .pau i'm looking forward(+1) to a trip(+1) to bamps(+1) and lake louis(+1) in a couple(+1) of weeks(+1) .pau .br .pau hope(+1) to do some biking(+1) .pau pedal(+1) biking .pau not .pau motorbiking(+1+1) .pau .ls .br .pau and(+1) i grew(+1) up .pau in the midwest(+1) so being(+1) out here in the northwest(+1) is a whole(+1) different experience(+1) for me .pau .br .pau i find(+1) the people to be very friendly(+1) .pau .br .pau i love(+1) the outdoor atmosphere(+2) .pau .br .pau i like(+1) being able to wear(+1)". (+1) = primary stress; (+2) = intermediate stress.

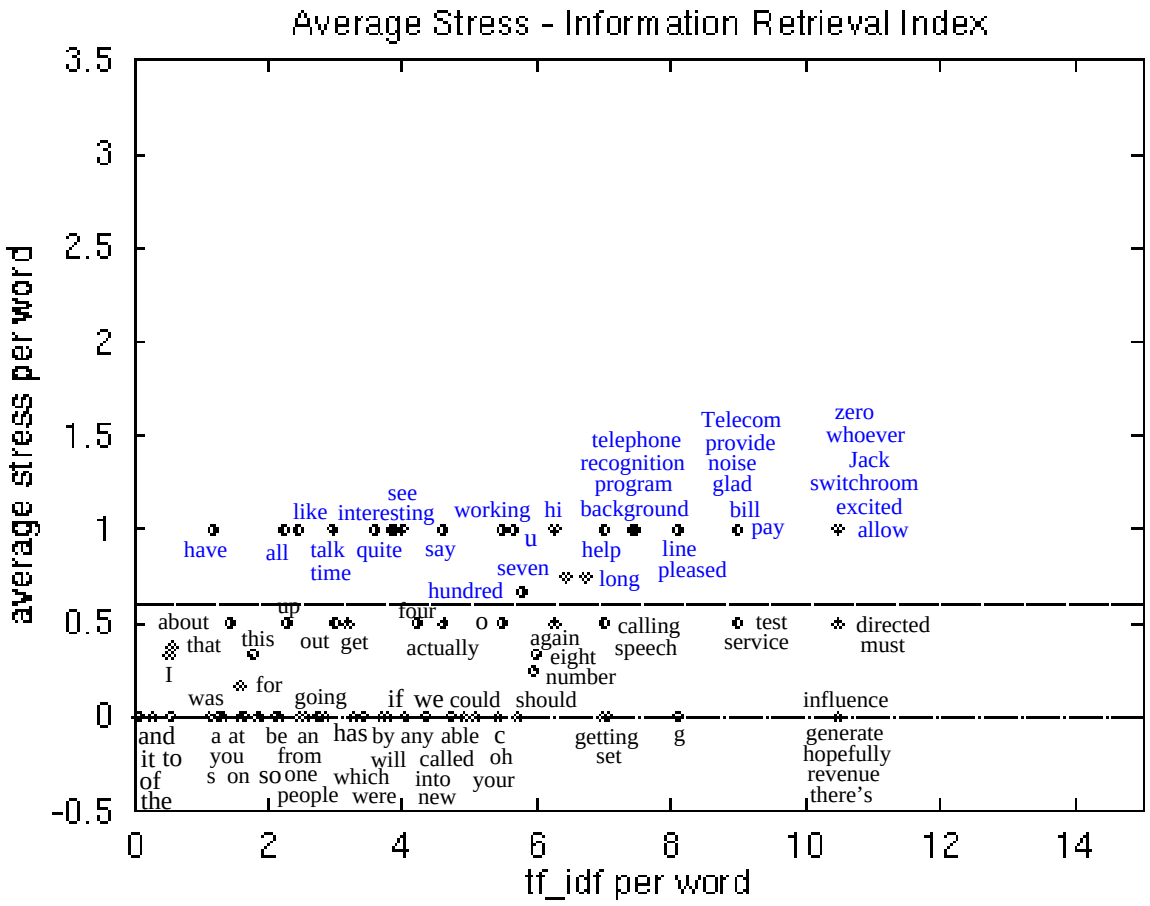

Figure 27: ".pau oh hi(+1) this is jack(+1) from o(+2) g c telecom(+1) actually(+2) i called(+1) out to test(+2) .pau out(+2) this new eight hundred(+1) number(+2) .pau .br to see(+1) if it was working(+1) which it is(+1) it's eight(+1) hundred four(+2) four(+2) one one zero(+1) seven(+2) seven(+1) .br .pau i must(+2) say(+1) i'm quite(+1) pleased(+1) as we uh provide(+1) your telephone(+1) service(+2) to be able to have(+1) an eight hundred(+1) number directed(+2) into a .pau program(+1) like(+1) this(+1) which uh is going to allow(+1) people to talk(+1) for a long(+1) long(+2) time(+1) .pau that'll generate lots(+1) of revenue for us(+1) and .br .pau hopefully uh .pau whoever(+1) .pau uh u(+1) s west(+2) will get(+2) to pay(+1) the bill(+1) for(+2) it .pau .br so were excited(+1) about(+2) that(+1) .pau again(+2) i'm calling(+2) from the switchroom(+1) so theres a lot of background(+1) noise(+1) which should be interesting(+1) to see(+1) .pau .br .pau if that has any influence at all(+1) on your uh your speech(+2) recognition(+1) .pau program(+1) .pau .br .pau uhm .pau i'm glad(+1) we could .pau .br .pau help(+1) .pau you out(+1) by getting that(+2)uh .br .pau getting that line(+1) set up(+2) .pau ". (+1) = primary stress; (+2) = intermediate stress.

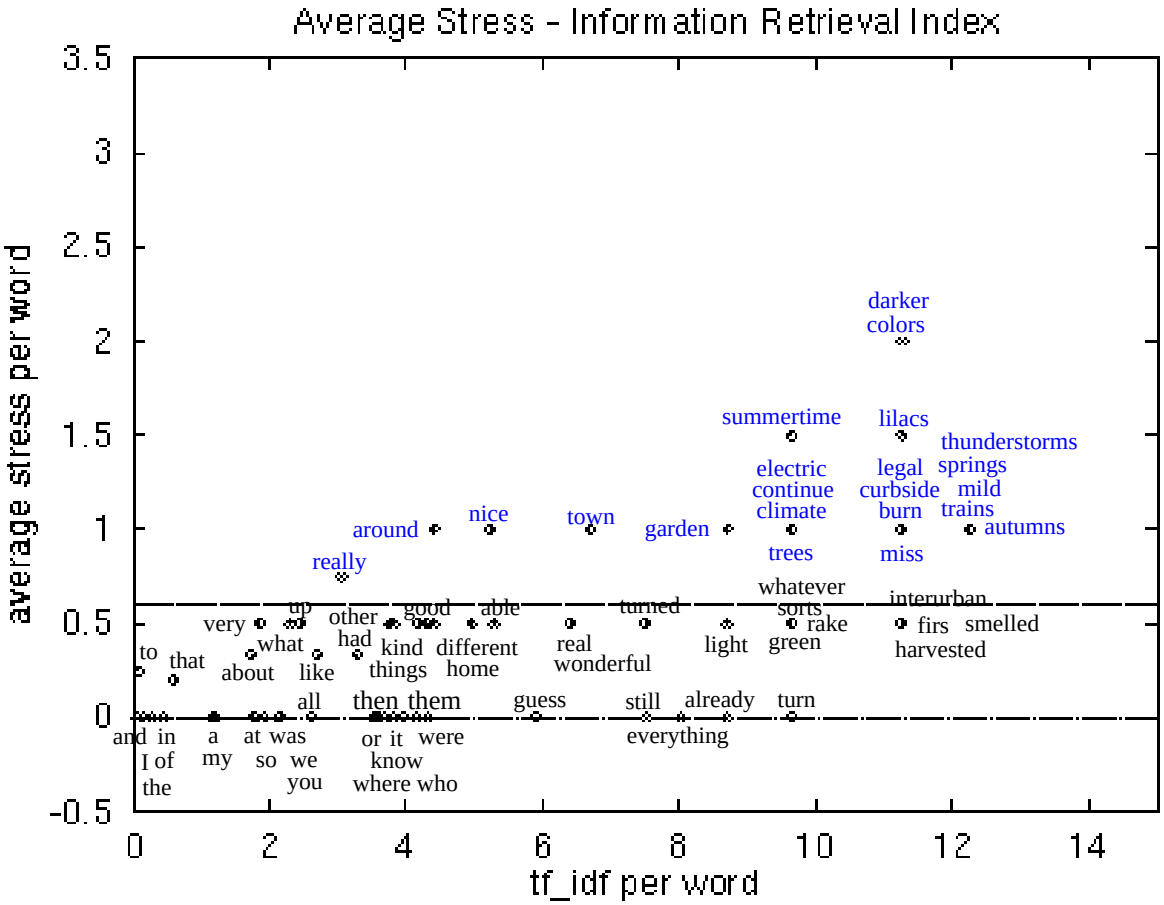

Figure 28: "guess I'll continue uh about what I like about the climate or whatever my home town uh I really like the thunderstorms in the summertime that was wonderful I miss that uhm .br there is very nice mild springs as things turn you know kind of light green first and then darker and it smelled so good and we had a lot of lilacs around .br in the autumns who were real autumns where the trees turned all sorts of different colors and we'd rake them up and we were still able to or where it was legal to burn them at the curbside or in the garden where you had already harvested everything .br that was really nice uhm the other things that I like about the town itself was that it had duh .br electric interurban trains".

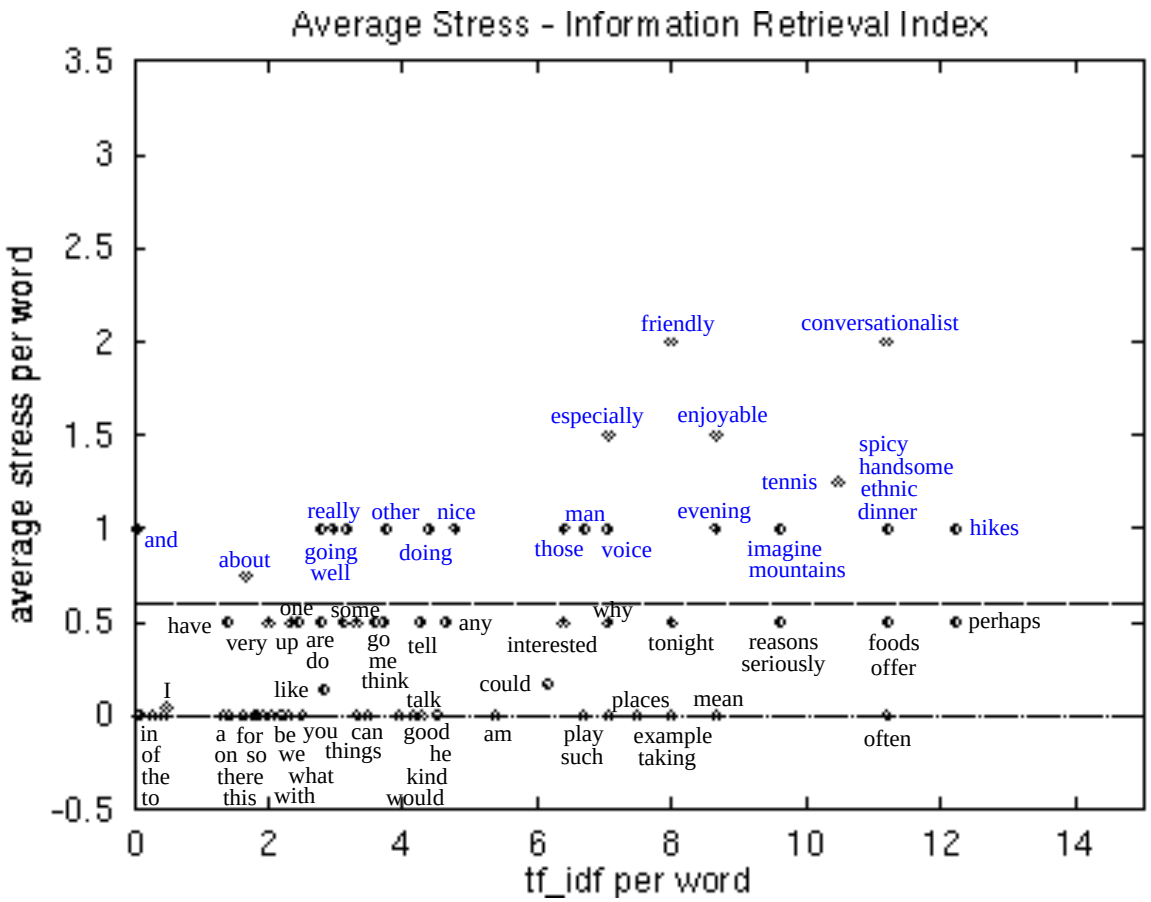

Figure 29: "Well I what I'd really like to talk about is what you are doing for dinner tonight you have such a nice voice .br uh I the kind of places I like to go are I like uh spicy foods like ethnic restaurants .br often have uh .br mean perhaps you'd be interested in uh going to one of those with me .br uh some evening .br uh I can tell you lots of reasons why I would be enjoyable I am a very good conversationalist .br I am friendly .br and I uh am handsome as any man you could imagine .br so perhaps you'll think seriously about taking me up on this offer .br there are other things we could do as well we could for example play tennis I like to play tennis I like to go for hikes especially .br uh he hikes in the mountains I like".

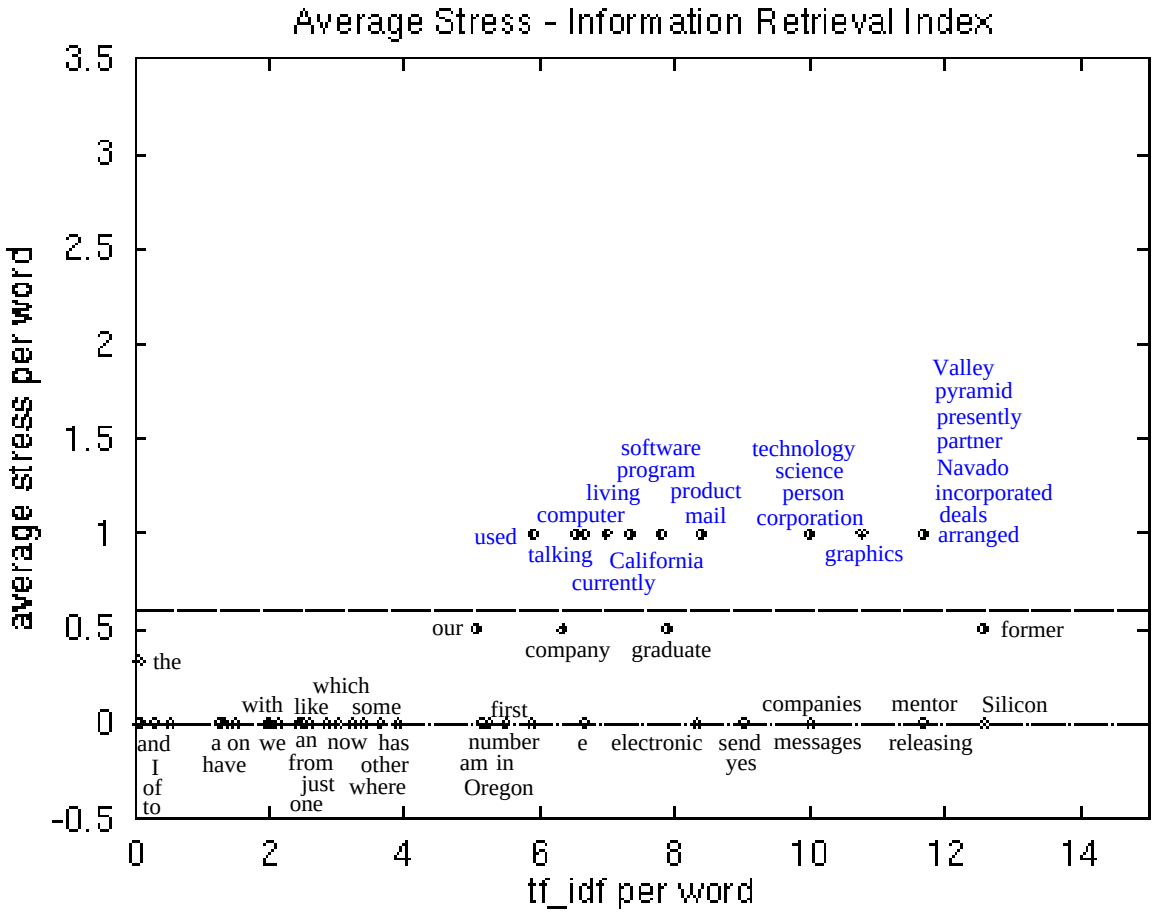

Figure 30: "I'm a former graduate of the former graduate yes I'm a graduate of the Oregon Graduate Institute of Science and Technology I'm presently living in uh .br Navado California where I uh am a partner in a software company .ls our company is currently a uh releasing a product and .br has just uh arranged some deals with mentor Graphics Corposration and uh .br uh Silicon Graphics Incorporated and uh Pyramid and a number of other uh computer companies in Silicon Valley .br .ls and we uh have as our first product an e mail program which is used to send electronic mail messages on computers like the one I'm now talking to .ls .br uh .br .ls from person to".

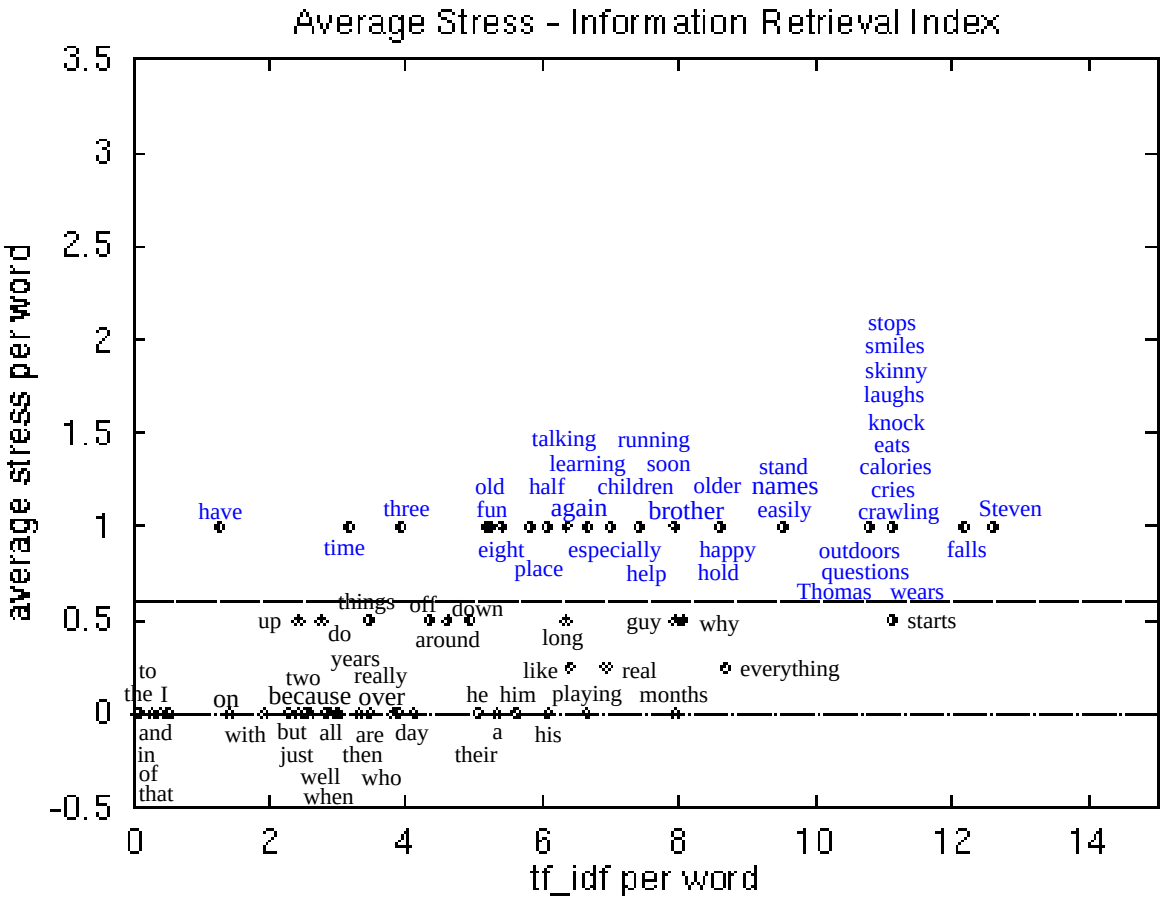

Figure 31: "Well I have two children and their names are Steven and Thomas Steven is three and a half years old and Thomas is eight months old .br they're lots of fun uhm Steven likes to do everything outdoors he's real skinny because he just wears off all the calories that he eats in running around all day long and talking and .br asking questions like why why why everything is why .br and Thomas is crawling all over the place and just learning to stand up and hold on to things .br and he falls down a lot .br especially with the help of his older brother who like to knock him down .ls but he's a really happy guy he cries when he falls down but then he stops real soon then starts playing again he laughs really easily and smiles all the time".

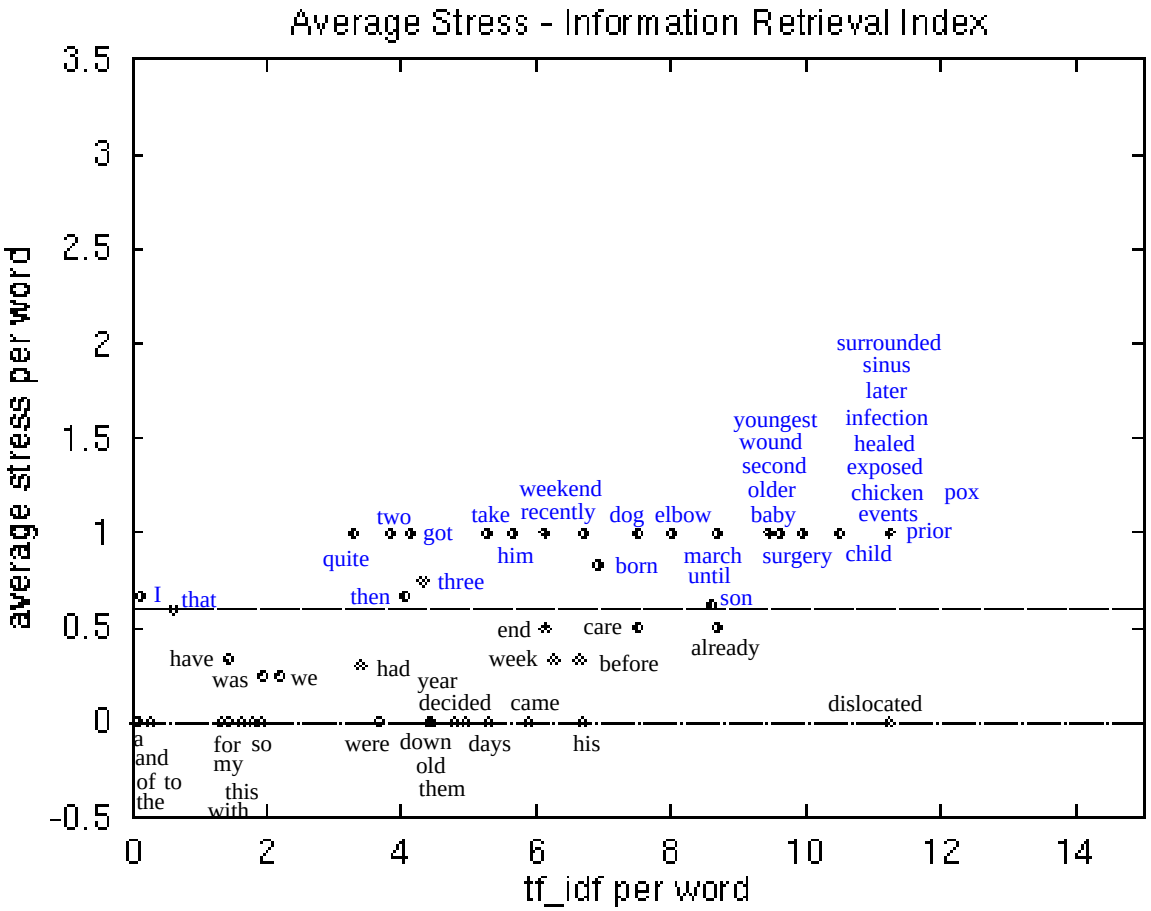

Figure 32: "Recently I had a baby .br uhm have already have a two year old son and decided to have a second child this the second child .br was born the end of march and we had lots of events that surrounded that .br three weeks prior to that my dog had surgery .br and we had to take care of him for two weeks until his wound was healed .br before that the weekend before that my older son dislocated his elbow .br and then I came down with a sinus infection the week the baby was born .br two days before the baby was born we were exposed to chicken pox and three weeks later .br I got them then my youngest son the baby got them and then .br my older so got them so we've had .br quite".

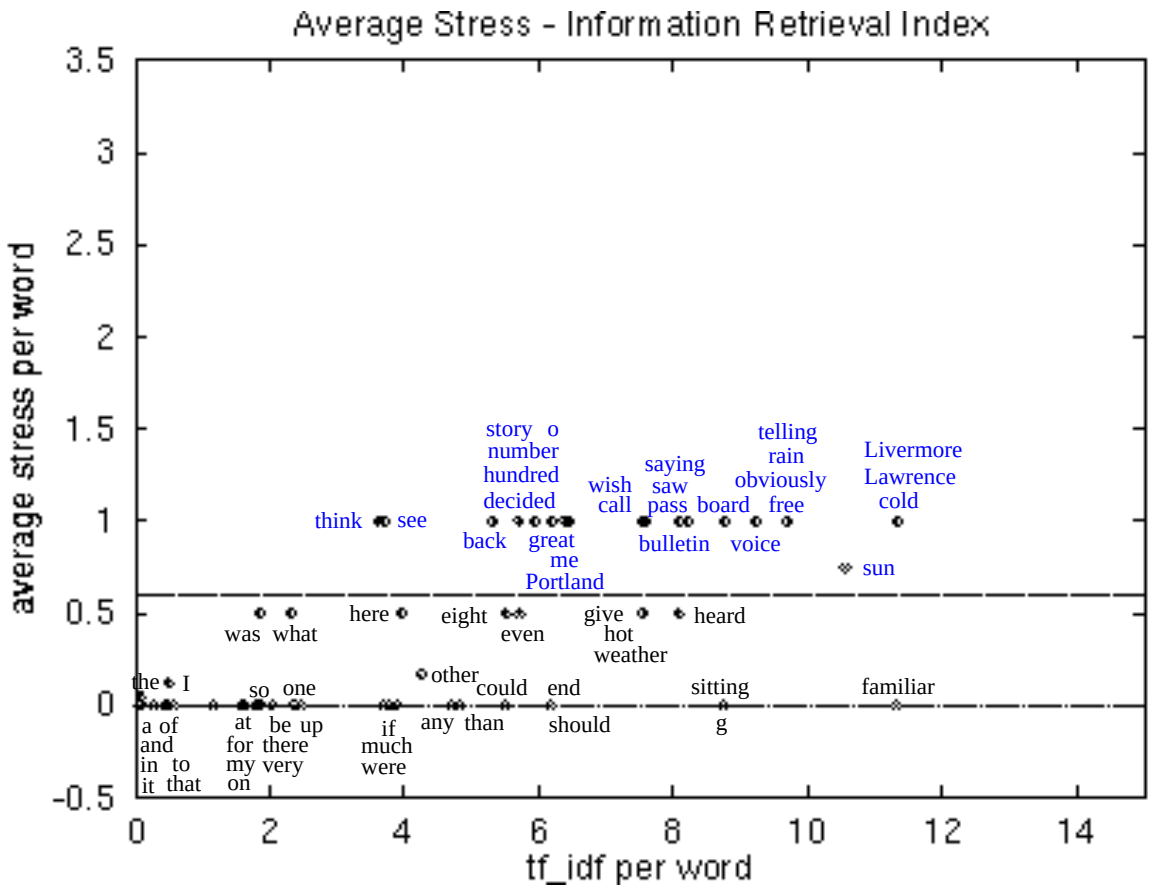

Figure 33: "Here I was sitting in the sun at Lawrence Livermore and I saw and a a .br message on the bulletin board and I saw that I could call OGI for free on a one eight hundred number .br so I couldn't pass that up .br .br and I decided to give it a call and see if there were any familiar voices on the other hand .br and I even heard Kelly's voice telling me .br what I should be saying for the story .ls obviously I can't think of very much other than .br other than and uh describing my call here the uh the weather here is great .br and that I wish I was back in the rain and the sun and the cold of Portland".

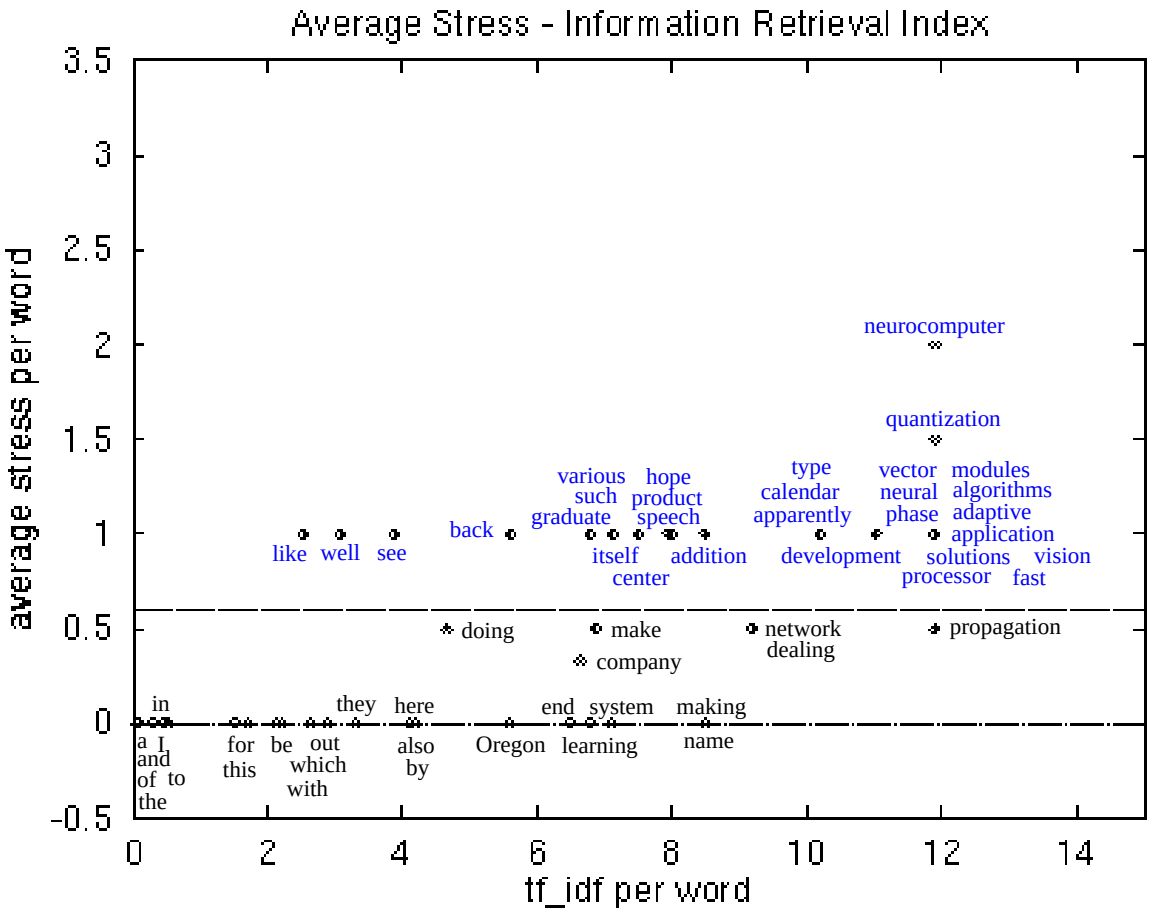

Figure 34: "Uh see here I'm associated with the company out here the Oregon Graduate center the name of the company is Adaptive Solutions .br uh the company makes uh .br uh neurocomputer which is a uh fast processor for doing neural network type algorithms .br such as uh back propagation uhm learning vector quantization and the like .br uh apparently in the development phase and uh hope to be uh .br making product by the end of this calendar here .br uh in addition to the development system itself they also make .br uhm .br various application modules dealing with uh .br aspects of vision and speech as well".

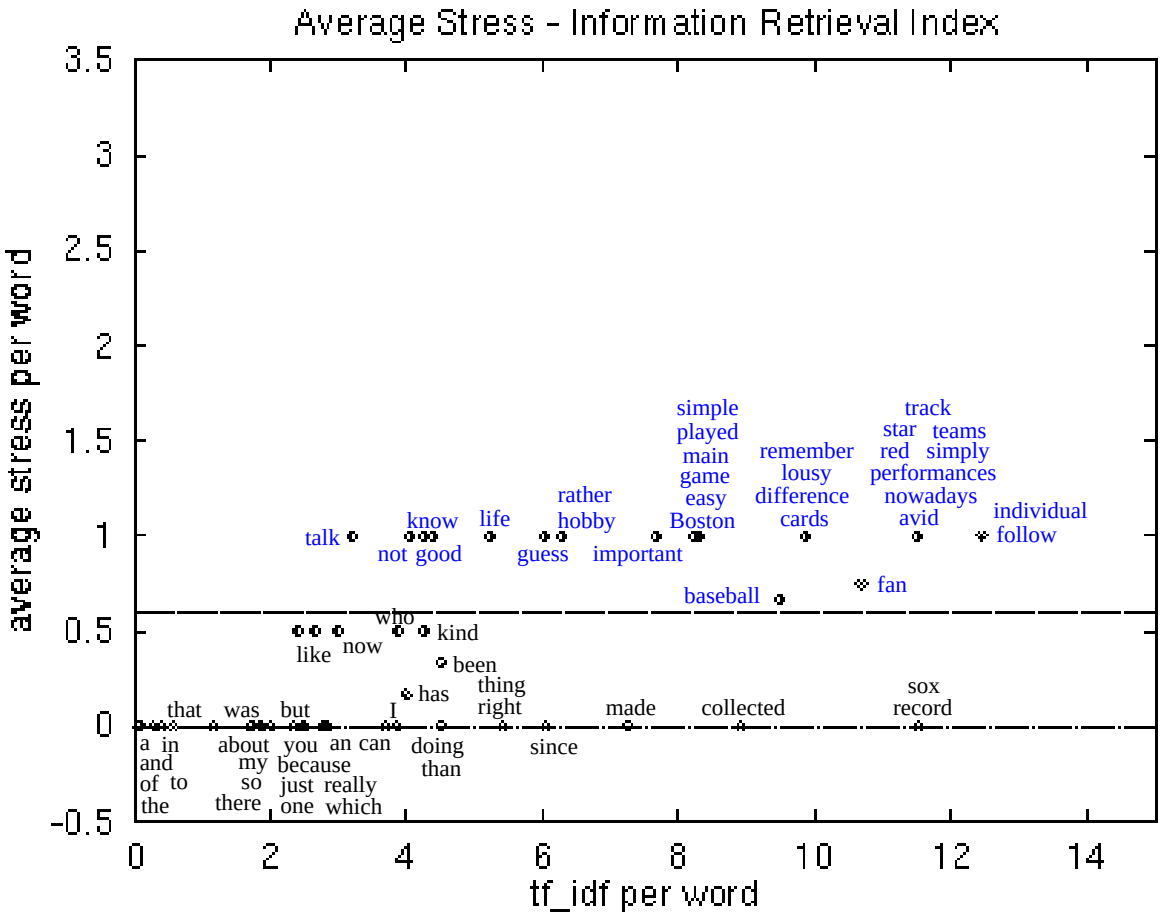

Figure 35: "I guess I'll talk about baseball which has been .br the main hobby of my life since I can remember .br uhm I've played baseball collected baseball cards .br .br uhm been an avid baseball fan .br uh right now I'm a fan of the Boston Red Sox which isn't so good nowadays since they're kind of doing lousy .br uhm one thing that I like about baseball is that .br you can follow individual performances rather than .br simply the teams record simple to know .br who .br has made a difference in the game and who has .br just been there .br and was important but but .br not really the star .br I I like baseball because it's easy to follow .br individual track records it's easy to".