COMPUTER-AIDED ANALYSIS OF KETAMINE-INFLUENCED SPEECH

by

CONGZHOU HE

(Under the Direction of Michael A. Covington)

ABSTRACT

This is a computer-aided analysis of the speech of healthy volunteers under the influence of ketamine as contrasted with their normal speech. The project was undertaken since ketamine has been shown to simulate some of the symptoms of schizophrenia and ketamine-influenced speech shares much similarity with schizophrenic speech. Thus, the study of ketamine-influenced speech provides information relevant to the diagnosis and treatment of schizophrenia.

I contrastively studied 20 pairs of normal and ketamine-influenced speech samples from 10 healthy volunteers at various linguistic levels: phonetic, syntactic, semantic and pragmatic. I conclude that ketamine-influenced speech exhibits several distinctive properties: a smaller percentage of verbs, a higher percentage of nouns, an increase in the use of fillers, and greater variability of vowel height.

I implemented software that recognizes the distinctive properties of ketamine-influenced speech. The software was tested on 37 pairs of testing data and achieved high accuracy in identifying ketamine-influenced speech.

INDEX WORDS: ketamine-influenced speech, computer-aided speech analysis, verbs, nouns, first formant, filled pauses

COMPUTER-AIDED ANALYSIS OF KETAMINE-INFLUENCED SPEECH

by

CONGZHOU HE

B.A. Shanghai International Studies University, China, 1992

M.A. Shanghai International Studies University, China, 1995

Ph.D. Shanghai International Studies University, China, 2000

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2004

COMPUTER-AIDED ANALYSIS OF KETAMINE-INFLUENCED SPEECH

by

CONGZHOU HE

Major Professor: Michael A. Covington

Committee: Donald Nute
Marlyse Baptista

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2004

ACKNOWLEDGEMENTS

I am deeply indebted to my major professor Dr Michael A. Covington for his keen insight and valuable advice. Without his consistent support, I would never have experienced the thrill and joy that resulted from this research.

I am very thankful to Dr Donald Nute and Dr Marlyse Baptista for their guidance as my committee members. I also wish to thank Dr W. D. Potter, my other professors, and my fellow students at the AI Center, the University of Georgia. Had it not been for their help, this thesis would not have been possible.

I am grateful to my dear husband Huimin Ji for discussing with me the theories behind my project and for giving me a comforting home that I can always return to. I thank my lovely baby daughter Yinyu Ji for showing me the joys of life through the hardship of thesis writing, and my parents who offer their generous assistance with the greatest patience.

This research was supported by a contract with GlaxoSmithKline Research & Development Ltd. ("Computer Analysis of Schizophrenic Speech," Dr Michael A. Covington, Principal Investigator). Special thanks go to Cati Brown who took on the hard job of transcription.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER ONE

INTRODUCTION

As human language closely reflects brain functions in a way that no other human activities do, linguistic analysis has been an integral part of psychiatric research. Many brain diseases such as schizophrenia exhibit language malfunctions that have been used in disease diagnosis. Recently, computer aided speech analysis has been much called for thanks to its speed, objectivity and precision. This thesis is an attempt at accurate computer aided analysis of slightly flawed drug-induced speech. Its goal is two-fold: to design and implement software that takes human speech data as input and outputs analyzable statistics of various measurements, and to use the software to study in depth the distinctive properties of ketamine-influenced speech and contrast it with normal speech.

1.1 What is Ketamine?

Ketamine is an anesthetic and analgesic invented by Calvin Stevens in 1962. Unlike most other anesthetics, ketamine is not a sedative in the usual sense of the word; rather, it has some unique cardiovascular-stimulating properties that accelerate heart rate and increase blood pressure. This has prevented it from being used as a general clinical anesthetic, yet it is also the reason why ketamine is among the more common and suitable anesthetics to be administered to children in oral surgeries and to some physically weak seniors (Jansen 2002; White, Way and Trevor 1982).

Ketamine features a rapid onset of effect, a prompt recovery and mild emergence reactions (White, Way and Trevor 1982: 126). “The effects begin about 30 seconds after an intravenous injection, 2 to 4 minutes after an intra-muscular injection, 5 to 10 minutes after snorting or intra nasal use and about 20 minutes after an oral dose on an empty stomach” (Jansen 2002). Because of this, ketamine is reported to have been successfully used on many burn patients for quick pain relief, who are also reported to have resumed nutritional intake after a rapid recovery from the ketamine used in the operation (White, Way and Trevor 1982: 126).

Like other anesthetics, ketamine exerts effects on the central nervous system (CNS) even with subanesthetic doses, and causes CNS changes that interfere with people’s “ability to organize thoughts and understand the environment during emergence” (White, Way and Trevor 1982: 119). This is partly achieved by affecting the CNS neurotransmitter systems and by “blocking the action of glutamate, which plays a crucial role in thinking, memory, motion, language, sensation and perception” (Jansen 2002). Recreational drug users are said to take “k-trips” with high doses of ketamine, where the psychedelic effects of ketamine cause near-death or near-birth experiences. K-trips can trigger neurosis by blunting perception of the outside world, suppressing the conscious and stirring up the unconscious (Jansen 2002; Meyer 1989).

Despite some unpleasant psychological experiences, ketamine is one of the safest anesthetics ever invented, considering the fact that its overdose hardly ever leads to death. The potential danger of using ketamine comes mostly from the physical weakness, numbness, and the inability to control one's own body in a hazardous environment (Jansen 2002). Both healthy and diseased volunteers have safely participated in numerous ketamine-related experiments using higher or lower doses, and their behavior and speech have been carefully recorded and examined to better human knowledge in the areas of anesthesiology, psychiatry and neurology.

1.2 Why Ketamine-Influenced Speech?

Ketamine has been widely studied and much is known about the drug. A simple keyword search of "ketamine" in the Medline database brings up more than 3,600 related papers. Among these is a small recent portion that examines the hypothesis that ketamine produces in healthy volunteers a set of schizophrenia-like symptoms (e.g., Malhotra et al. 1996; Adler et al. 1998; Lahti et al. 2001). Ketamine takers report experiences such as hearing sounds, seeing God, and lucid dreaming, which very much simulate what schizophrenic patients undergo. Experiments relating ketamine with schizophrenic symptoms have been carried out mainly through direct observation of the volunteers' behavior and, commonly, by comparative study of the volunteers' speech produced in the course of ketamine administration. These studies provide not only a detailed psychosis model on the behavioral effects of ketamine, but also reasonable suggestions as to the origins, developments and treatments of schizophrenia –

“a chronic, severe and disabling brain disease” that affects about 1% of the human population (National Institute of Mental Health 1999).

Since speech mirrors thoughts consciously or unconsciously, and since speech is directly observable, diagnoses of schizophrenia are frequently partially based on a patient’s abnormal use of language. In fact, all of the three major diagnostic classification systems for schizophrenia, i.e., ICD-10 (WHO 1992), Diagnostic and Statistical Manual, fourth edition (DSM-IV) (American Psychiatric Association 1994), Research Diagnostic Criteria (RDC) (Spitzer, Endicott and Robins 1978), heavily rely on various speech disorders for diagnostic criteria. As was mentioned in 1.1, ketamine blocks glutamatergic receptors, which are crucially involved in memory and thinking. Among these receptors is the N-methyl-d-aspartate (NMDA) receptor frequently implicated in the pathophysiology of schizophrenia (Malhotra et al. 1997: 141). The study of ketamine-influenced speech can, therefore, have significant implications as to what exactly are the distinctive properties that make schizophrenic speech abnormal, and thereby help in early and accurate diagnosis of schizophrenia.

Researchers have studied the effects of ketamine on the behavior of healthy volunteers (e.g., Adler et al. 1998; Malhotra et al. 1996; Lahti et al. 2001). They have agreed on the fact that subanesthetic doses of ketamine induce linguistic characteristics that have been used as diagnostic criteria for schizophrenia. The characteristics that have been repeatedly reported in the literature include:

1) Disorganized speech and derailment: Ketamine-influenced speech is reported to be loosely associated, featuring flight of ideas and disconnection of thoughts.
2) Incoherence and illogicality: “Bizarre reasoning” (Lahti et al. 2001: 463) is often found in ketamine-influenced speech.
3) Emotional-withdrawal and poverty of speech: Ketamine significantly affects verbal productivity.

Abnormal speech, however, is only a partial focus of the studies on ketamine-induced behavior, and hence there are only some brief descriptions of the above characteristics without further details.

This thesis, on the other hand, aims at a closer look at the linguistic details that account for the abnormality felt by the researchers in previous studies and which can be objectively measured and verified in the future.

1.3 Why Not Existing Psychiatric Scales?

One important goal of the thesis is to devise some accurate measures for speech abnormality, and for a computer to use the measures to distinguish ketamine-influenced from normal speech. In the course of achieving this goal, I decided not to directly use the existing psychiatric scales or selective items from these scales, which have been widely adopted in previous studies.

As mentioned in 1.2, significant statistics have been successfully obtained to find some distinctive characteristics of ketamine-influenced speech. The statistics have been acquired based on some well-established scales that are very frequently used in schizophrenia related research. Some of these scales are:

1) Brief Psychiatric Rating Scale (BPRS): BPRS has been the most popular psychiatric rating scale for more than 30 years with proven high validity, especially in the study of schizophrenia. Initially published in 1962 (Overall and Gorham 1962), it has now developed to be a 7-point scale on 24 items. Among these, more than 10 items are rated primarily on the basis of verbal responses, such as conceptual disorganization, unusual thought content, blunted affect and hallucinatory behavior (Overall and Gorham 1962; Van Riezen and Vrijmoed-de Vries 2000; American Psychiatric Association 2000).
2) Scale for the Assessment of Thought, Language and Communication (TLC): In her widely-cited paper "Thought, Language, and Communication Disorders" (Andreasen 1979a), Andreasen used 18 items to define "formal thought disorder" or "disorganized speech." All of the items are speech related, with fairly precise descriptions for each to increase the reliability of the scale. For instance, "neologism" is clearly defined as "a completely new word or phrase whose derivation cannot be understood," and "self-reference" as "a disorder in which the patient repeatedly refers the subject under discussion back to himself when someone else is talking and also refers apparently neutral objects to himself when he himself is talking" (Andreasen 1979a).
3) Scale for the Assessment of Positive Symptoms (SAPS): SAPS (Andreasen 1984) is used to assess the severity of positive symptoms mainly in schizophrenia. Its 5 main categories (i.e., hallucination, delusions, bizarre behavior, positive thought disorder, inappropriate

affect) include 31 items. Most of the speech related items are adopted from TLC (American Psychiatric Association 2000).

4) Scale for the Assessment of Negative Symptoms (SANS): Complementary to SAPS, SANS (Andreasen 1982) is designed to rate the severity of negative symptoms mainly in schizophrenia. It consists of 19 items that fall into 5 categories: affective flattening or blunting, alogia, avolition-apathy, anhedonia-asociality, and attention. Some of the important speech related items are: lack of vocal inflections (i.e., flat intonation), poverty of speech, poverty of content of speech, blocking, and increased latency of response (American Psychiatric Association 2000).

While the above mentioned scales outline some significant features of speech disorders from mental diseases, they and their speech related items are unsuited to the current study as they are for the following reasons:

1) These psychiatric scales are designed for behavioral abnormality and not speech abnormality specifically. Hence, they are clinical interview based and require keen observation of paralinguistic features such as eye contact and facial expression as well as linguistic features. This study, however, is based solely on audio recordings, which precludes the possibility of examining paralinguistic features. Even for the linguistic items in these scales, many of them, especially those related to speech content and logic, are left hard to rate, since the subjects in my project are required to describe pictures rather than conduct conversations with the interviewers.

2) The items in the existing scales, however well defined, are based on abstract standards and subjective impressions. Because of this, in order to achieve desirable reliability, different versions of the scales have been developed "with detailed anchor descriptors and semistructured interviews" (American Psychiatric Association 2000: 493), and interviewers need considerable training. For my project, in contrast, the computer requires explicit directions on actions to take on digitized data to actually score speech samples and reach a conclusion objectively. For any item in the psychiatric scales to be useful in this project, it needs to be concretized to be a computable measure of tangible data that may be extracted and analyzed automatically.

3) The existing scales are more suitable for detecting moderate to severe levels of psychopathology than mild levels. Subtle changes in intonation, for instance, may not be so substantial as to result in affective flattening that can be captured by the interviewer. Slight changes in the choice of vocabulary are even more elusive to the human audience, if the speech does follow the flow of logic. In addition, certain items such as neologism and clanging are of little or no use to this project, since the occurrences of these language phenomena are uncommon even for patients with severe mental disorder. More finely defined and more typical features that can be objectively measured should be devised for the purposes of this study.

4) Besides physical variances of the individual subjects, differences in regimens and recording times severely complicate experiment results. As mentioned in 1.1, ketamine has a faster onset and recovery than most other anesthetics. Previous research has proven that

differences in the ratings by the psychiatric scales stay significant only very briefly, and fade after about 45 minutes after ketamine. Many studies (e.g., Adler 1998; Lahti 2001) agree that there is no significant effect at 90 minutes after ketamine. Yet, there is no consensus as to at what time the effects of ketamine peak. For instance, Lahti (1995, 2001) suggests 20 minutes after injection while Adler (1998) proposes 30-45 minutes. To avoid complication, this project makes no specific requirement as to the timing of the recording other than the fact that ketamine should still be in effect. Moreover, the regimens in published studies are slightly higher than the regimens in this project. It is therefore imperative that the linguistic measures to be used in the project should undergo significant changes under the influence of ketamine without being too time- and dose- sensitive.

5) A major goal of this project is to discover aspects of ketamine-influenced speech that have not already been recognized or published. These aspects may include subtle changes that easily elude the human ear and were, hence, not mentioned in previous studies, but which together bring about the abnormality that distinguishes ketamine-influenced from normal speech.

Despite what has been said, the existing psychiatric scales, especially their speech related items, provide meaningful guidance for the design of the current study as will be discussed soon.

1.4 Current Study at a Glance

As was mentioned, this is a computer aided study of ketamine-influenced speech. The following three chapters will each discuss one of the three stages of the study:

In Chapter Two, I design and implement a software package named Ketamine Speech Analysis Software (KSAS) that extracts numerical data from digitized audio recordings of human speech. KSAS outputs over 130 statistics that are designed to detect speech disorders, and its usefulness is not restricted to the identification of ketamine-influenced speech.

Equipped with KSAS, I compare in Chapter Three the normal and ketamine-influenced speech samples of ten healthy volunteers. The speech samples were produced when the volunteers were requested to describe the pictures from the Thematic Apperception Test (TAT, Murray 1943/1971). I use statistical methods to find out language properties that are significantly influenced by ketamine regardless of recording time and dosage.

In Chapter Four, I design a simple but accurate algorithm to incorporate the important measures found in Chapter Three. The algorithm distinguishes normal from ketamine-influenced speech when both speech samples of the same individual are input to KSAS. I have tested the algorithm on 37 pairs of speech samples and achieved high accuracy.

The final chapter concludes the thesis with some plans for future work.

CHAPTER TWO

KSAS – A TOOL FOR SPEECH ANALYSIS

The first stage of the project is to develop a software tool for speech analysis. A software package named Ketamine Speech Analysis Software (KSAS) has been implemented, whose inputs, user interface and outputs will be discussed in this chapter.

2.1 The Inputs

Assuming that all the audio recordings of speech samples can be digitized and saved in the form of .wav files, the first task is to decide what linguistic information needs to be obtained from the .wav files so that it can be further analyzed. Aiming at a relatively thorough study of human speech, I shall examine speech samples from the perspectives of phonetics, semantics, syntax and pragmatics. The research in the latter three aspects requires close transcription of the sound files.

As for phonetic analysis, the essential information needed is retained if the data for amplitude, voicing, the fundamental frequency (i.e., F0), and the first four formants (i.e., F1, F2, F3, and F4) can be acquired from the .wav files. To obtain these data, the open source speech research tool WaveSurfer (Center for Speech Technology 2004) is used to preprocess the .wav files of the speech samples.

WaveSurfer incorporates the functions get_f0, for automatic pitch extraction, and formant, for automatic formant tracking, from the Entropic Signal Processing Suite (ESPS) of speech analysis tools, which was licensed in 2002 to the Centre for Speech Technology by Microsoft and AT&T (Kåre 2002). The pitch and formant tracking algorithms implemented in ESPS were introduced in Talkin (1995), Secrest and Doddington (1983), McCandless (1974), and elsewhere. Figure 2.1 is a snapshot of WaveSurfer performing speech analysis on a .wav speech sample.

WaveSurfer can output two types of files for speech analysis: .f0 and .frm files, both of which are text files each containing a table of numbers with each row recording important speech data every 10 milliseconds. The .f0 files contain four fields: F0, voicing (1 for voiced and 0 for voiceless), amplitude, and cross-correlation (a measure of regularity in the waveform, i.e., sonority). The .frm files represent formant information in eight fields, the first four of which are the frequencies of the first formant (F1), the second formant (F2), the third formant (F3), and the fourth formant (F4).

Hence, for each speech sample, the inputs to KSAS will be a set of three text files derived from one .wav file: the .f0 file (generated by WaveSurfer), the .frm file (generated by WaveSurfer), and the .txt transcription file (made by a human typist). These files together preserve essential linguistic information at all levels and provide the basis for further statistical analysis.

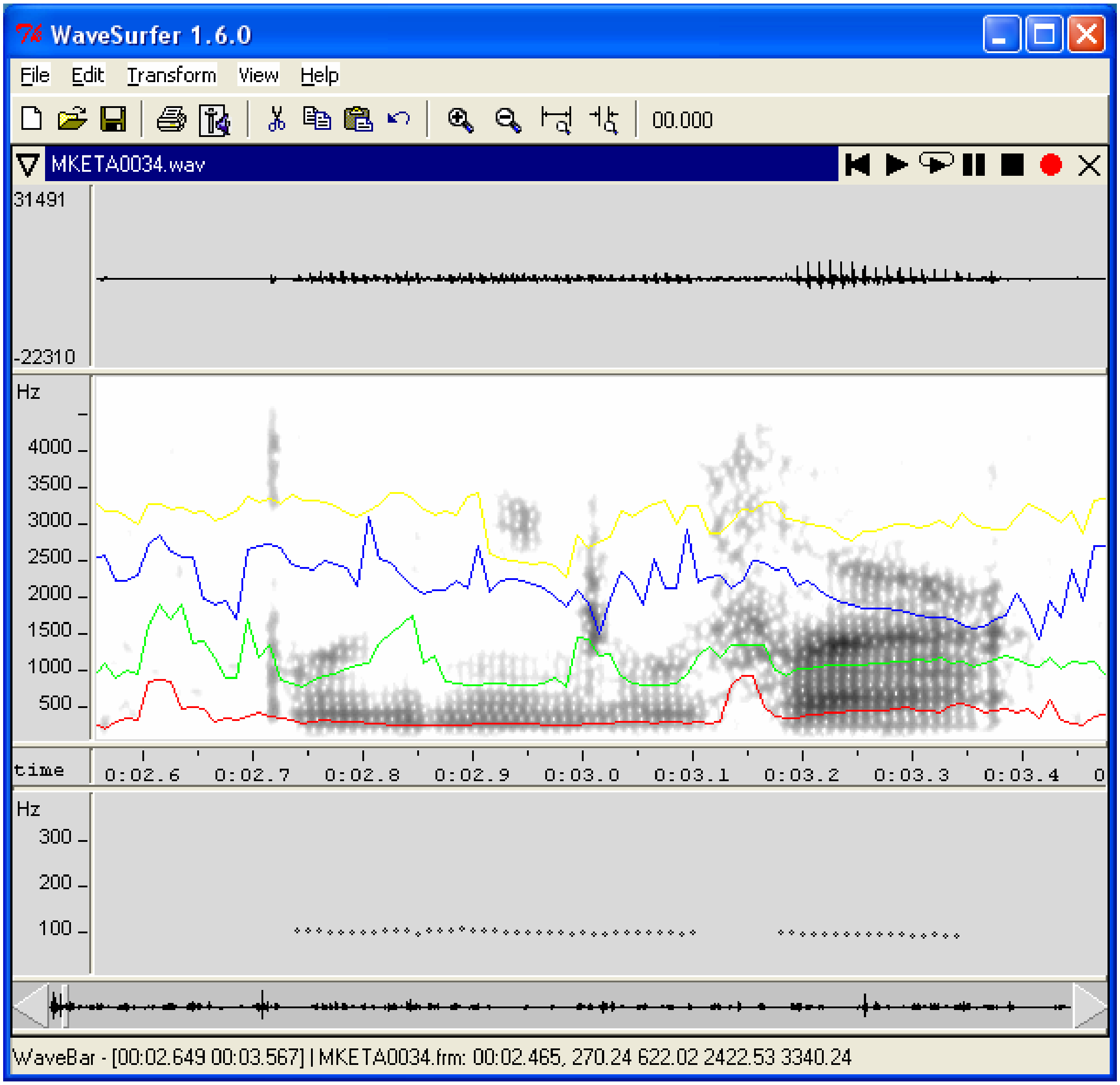


Figure 2.1 A snapshot of WaveSurfer speech analysis functions

It is crucial to understand that for my project, each .wav file should be the recording of exactly one speaker so that the subsequent computation can produce correct statistics for this particular speaker. For speech samples that contain the voice of an interviewer, editing of the .wav files is necessary to remove the interviewer's voice.

2.2 The Interface

KSAS has a user-friendly interface that was completely implemented in Java. Figure 2.2 is the main window for the user interface, where most of the functions can be performed either by using the pull-down menus or the buttons at the top of the window.

Since KSAS is designed to distinguish ketamine-influenced from normal speech, it accepts two speech samples of the same speaker. One and only one of the two speech samples is produced under the influence of ketamine. For each of the speech samples, there is a set of three input text boxes, where the names of the .f0, .frm, and .txt transcription files are displayed along with their paths when they are opened. I chose to display only the file names, because once created, these input files do not normally need editing, and because the .f0 and .frm files, which normally contain tens of thousands of lines of data, are too large to display in the main window.

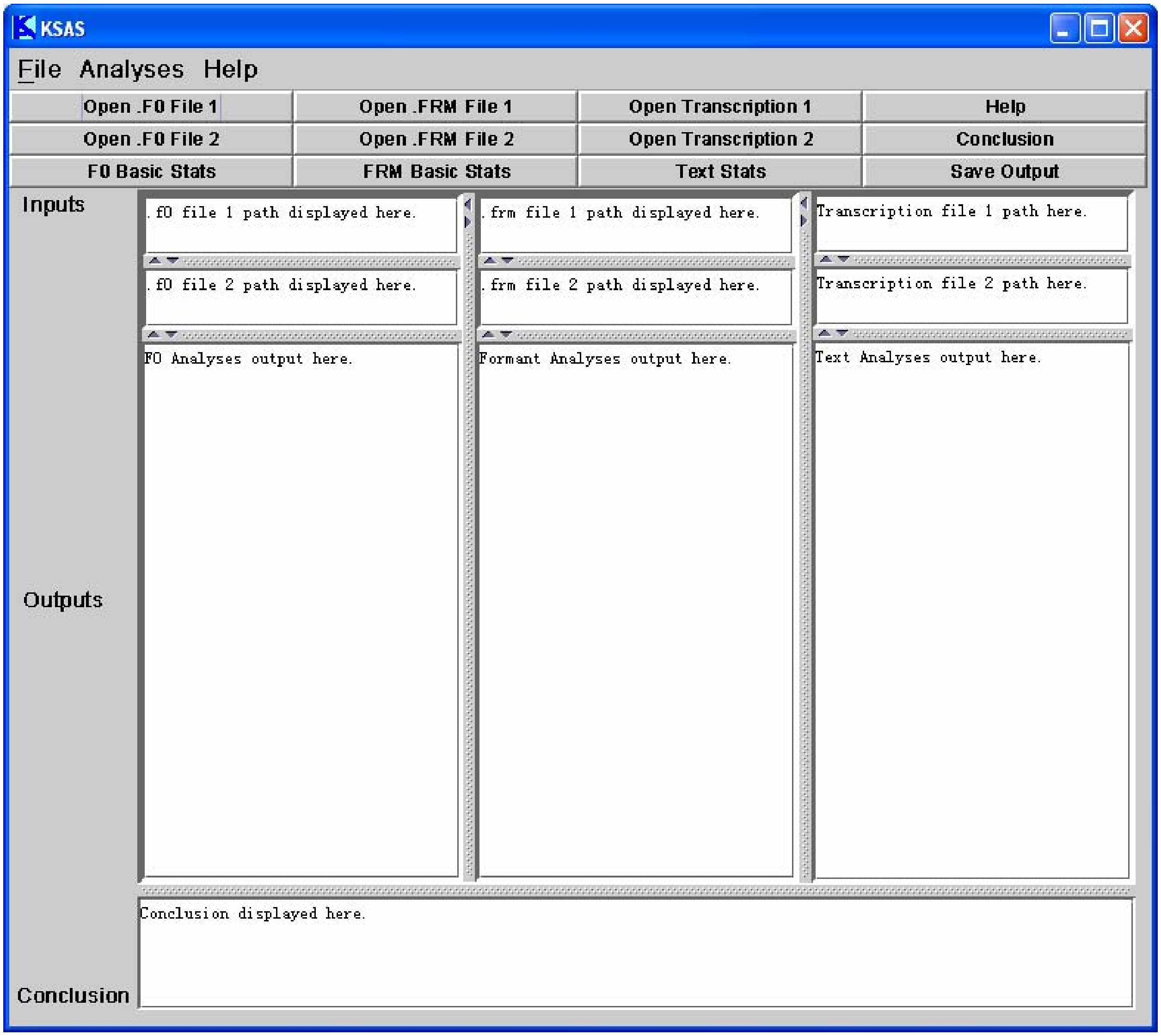


Figure 2.2 The user interface of KSAS

On opening all the necessary input files, the user can do F0 analyses, FRM analyses and text analyses on the .f0, .frm and .txt files respectively. Many of the measures are computed using data from more than one of the three files. For example, to compute the number of words uttered per second, text information contained in the transcription file should be combined with time information in the .f0 file. Therefore, it is required that the analyses be done in sequence. The outputs of various statistics will be displayed in the text boxes under the corresponding input files and will be further discussed in 2.3.

With the statistics from the speech analyses, KSAS will be able to draw a conclusion with high accuracy as to which of the two input speech samples is normal and which is produced under the influence of ketamine, and present it inside the conclusion text box at the bottom of the main window. The algorithm used to identify ketamine-influenced speech will be discussed in Chapter Four. The “Save Output” function saves everything currently in the text boxes into a .txt file when the button is clicked. Figure 2.3 is a snapshot of a sample run of KSAS. For a detailed KSAS user guide, please refer to Appendix A.

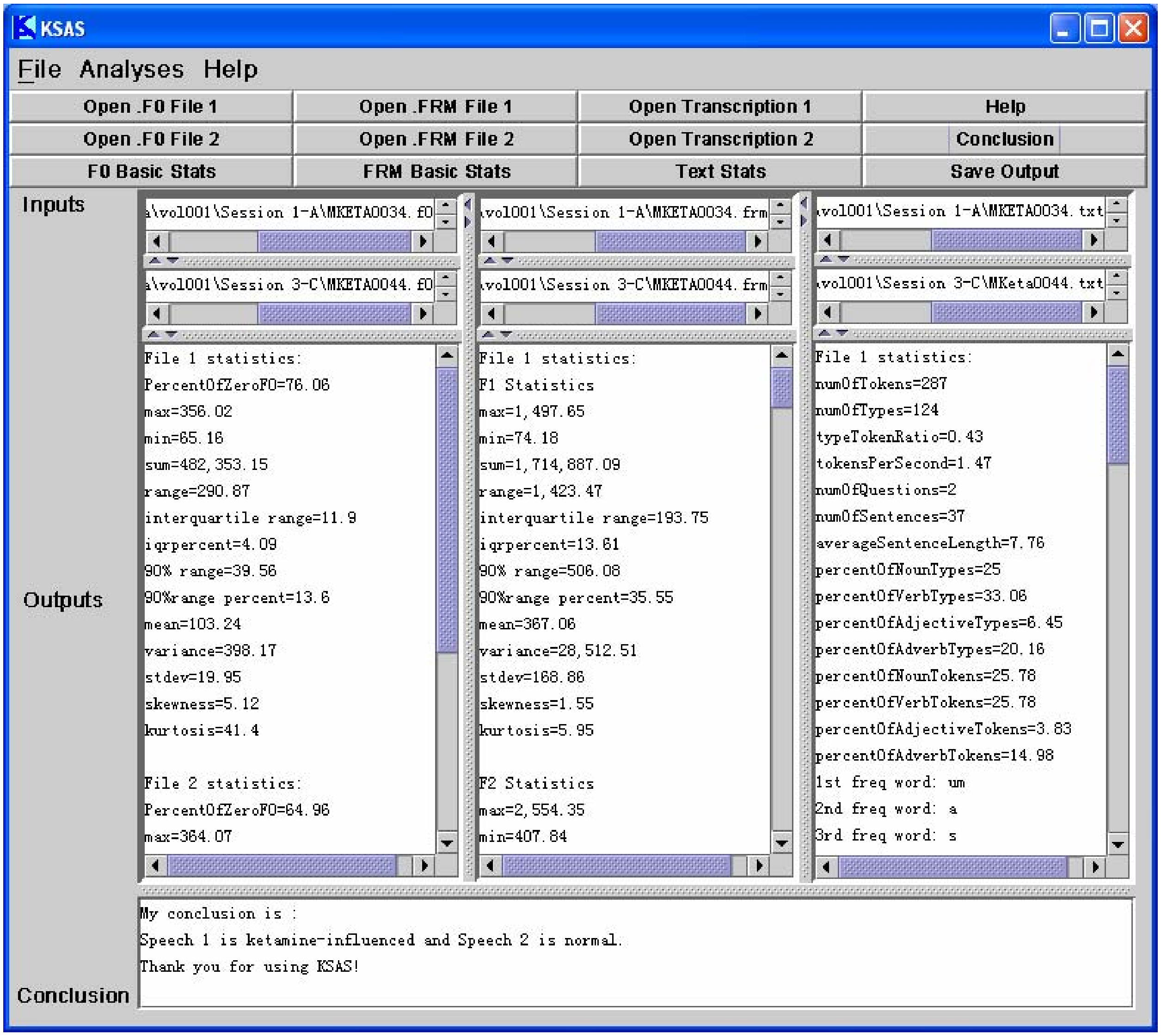


Figure 2.3 A snapshot of a sample run of KSAS

2.3 The Outputs

KSAS is first of all a software package to assist the analysis of all sorts of human speech by extracting useful statistics. Because of this, the program outputs over 130 statistics from the analyses rather than merely the 10 items that will actually be used for the conclusion regarding the effects of ketamine.

Below is a list of the statistics output by KSAS:

1) Percentage of zero F0s: When F0 is zero, there is no detectable vocal cord vibration in the speech sample. Percentage of zero F0s is a reliable indicator of pauses in the speech and speech fluency, and hence is a significant statistic for mental disorder with negative symptoms.

2) Basic statistics for F0: I use the term "basic statistics" to refer to a set of 13 measures of data distribution: maximum, minimum, sum, range, interquartile range, interquartile range as a percentage of the total range, range between the 5^{th} and the 95^{th} percentile, the range between the 5^{th} and the 95^{th} percentile as a percentage of the total range, mean, variance, standard deviation, skewness, and kurtosis. Some of these are major measures of central tendency and variation; others such as skewness and kurtosis provide more subtle information about the shape of the curve. Especially useful are measures like interquartile range as a percentage of the total range, which effectively reduces the influence of outliers in the data.

   The fundamental frequency (F0) is one of the most studied topics in acoustic phonetics, since its variation is manifested by the change in intonation, or the tone of an utterance (Ladefoged 2001: 192). And tone, in turn, is a crucial suprasegmental feature that reflects the speaker's mental state such as emotion. Hence, statistics for F0 can be particularly useful in psycholinguistic and psychiatric research.

3) Basic statistics for F1, F2, F3, F4, F2–F1, F3–F2 and F4–F3: I apply the same set of 13 measures to each of the first four formants as to F0. These formants are useful indicators of sound quality and reflect facts about place and manner of articulation. The frequency of the first formant, for instance, "shows the relative vowel height quite accurately" (Ladefoged 2001: 193). The differences between the formants are no less important than the formants themselves in describing the quality and position of speech sounds. For example, the backness of a vowel "is best related to the difference between the first and the second formant frequencies" (Ladefoged 2001: 177). Changes in the place and manner of articulation often result from neurological diseases, which alter a patient's speech from normal in various ways. Variations in formants along with those in the fundamental frequency are the focus of many studies in the past three decades (e.g., Haley, Ohde and Wertz 2001; Stassen et al. 1995; Kent 1979; Tolkmitt 1982; Stein 1993).

4) Text statistics: These measures are designed to describe the transcription text from the perspective of semantics, syntax and pragmatics. Most of the semantic measures concern lexical retrieval and comprise standard items such as: number of tokens, number of types, and type token ratio, which is a measure of repetitiveness vs. introduction of new ideas. Another set of eight unconventional measures concerns the percentage of various parts of speech used in the text. They are: percentage of noun types, percentage of verb types,

percentage of adjective types, percentage of adverb types, percentage of noun tokens, percentage of verb tokens, percentage of adjective tokens and percentage of adverb tokens. The data are retrieved by looking up words in the widely used electronic dictionary WordNet, developed by the Cognitive Science Laboratory at Princeton University (Miller et al. 2003). To access WordNet, KSAS calls the JWNL, which is an open source API developed at Stanford University for accessing WordNet-style relational dictionaries (Didion et al. 2003). Value changes in the semantic measures can partially reflect the impaired lexical retrieval and abnormally organized semantic network often reported in patients of mental diseases (e.g., Allen, Liddle and Frith 1993; Chance et al. 2002).

The speech of the mentally ill often displays simplified syntax, such as shallower clausal embedding and fewer relative clauses (Morice and Ingram 1982; Morice and McNicol 1985, 1986). KSAS computes four syntactic measures: the number of sentences, average sentence length for syntactic complexity, the number of occurrences of *that* per 100 words for embedded clause and for indexical uses, and the number of occurrences of *it*, *he*, *she*, *they*, *him*, *her* and *them* per 100 words for pronoun usage. These measures are effective although somewhat naïve without a built-in syntactic parser.

To machines, pragmatic changes are less tangible than changes at any other layer of language. KSAS focuses itself on conversational skills. The pragmatic measures used are: the number of questions (talking to the interviewer or oneself), the frequency of fillers like *um*, *eh*, *uh* and *oh*, the number of occurrences of the word *something* per 100 words, the frequency of negations as the sum of the number of occurrences of *no* and *not* per 100 words, the frequency of positive assertions as the sum of the number of occurrences of *yes*, *yeah* and *yep* per 100 words, the number of occurrences of the pronoun *you* per 100 words, and the number of occurrences of the pronouns *I* and *me* per 100 words. These measures roughly show the speaker's interpersonal communicative skills, his use of filled pauses and vague language, and some other conversational characteristics.

Through the above mentioned measures, KSAS not only provides a relatively complete description of a speech sample, but also concretizes many of the linguistic measures adopted in the psychiatric scales mentioned in 1.3. "Self-reference" used in TLC, for instance, is directly related to the frequency of the pronouns *I* and *me*. These measures can easily be adopted in other psycholinguistic investigations. In the next chapter, we shall see how these statistics have actually been used in the comparative study of ketamine-influenced and normal speech.

CHAPTER THREE

KETAMINE-INFLUENCED VS. NORMAL SPEECH

Equipped with the speech analyses software that I developed in Chapter Two, I will closely examine some speech samples and research the differences between ketamine-influenced and normal speech. In this chapter, I shall present some statistical analyses on the data extracted by KSAS, and then discuss the significant properties of ketamine-influenced speech uncovered by the analyses.

3.1 Data Extraction

The sound files I use for this project are recordings which were made during a placebo-controlled, double-blind study on the effects of ketamine conducted at Cambridge University, and which were provided to the University of Georgia by GlaxoSmithKline Research & Development Ltd. under a research contract. (Informed consent was obtained, and institutional human-subjects approval was obtained at both universities.) Twenty healthy volunteers participated in the study, each of whom was administered one of the two sets of three different subanesthetic regimens intravenously in random order. Set 1 includes the following regimens:

A) A 10-minute-long 0.27mg/kg bolus of ketamine followed by a 90-minute-long

0.14mg/kg/hr ketamine infusion;

B) A 10-minute-long 0.08mg/kg bolus of ketamine followed by a 90-minute-long 0.04mg/kg/hr ketamine infusion;

C) Placebo.

Set 2 includes the following regimens:

A) A 5-minute-long 0.26mg/kg bolus of ketamine followed by a 95-minute-long 0.97mg/kg/hr ketamine infusion;

B) A 10-minute-long 0.40mg/kg bolus of ketamine followed by a 90-minute-long 0.21mg/kg/hr ketamine infusion;

C) Placebo.

The doses of ketamine are slightly lower here than in most published studies; yet, as was mentioned in 1.3, the purpose of this study is to find distinctive properties of ketamine-influenced speech that are not too dosage-dependent.

The volunteers were asked to describe four pictures from the Thematic Apperception Test (TAT, Murray 1943/1971) as fully as possible during each of the three regimens. Seventeen volunteers had all their three recordings well saved (i.e., two after ketamine and one after placebo) and the other three had one of their ketamine-influenced files corrupted, leaving only the other two analyzable audio files (i.e., one after ketamine and one after placebo). Hence, we have 37 contrastive pairs of normal and ketamine-influenced (regardless of high or low dose) speech samples in all. Hardly any of the speech samples are distinguishable by the human ear as to whether they are produced under the influence of ketamine or not, except for one or two that are marked by exceptionally long pauses. The contents of the speech samples were logical and the speakers concentrated on the description of the pictures, showing no sign of digression. The speech samples had correct morphology, with proper inflections. They were syntactically acceptable, although packed with features such as hesitations, false starts, fillers, hedges, repairs and repetitions that are typical of any spontaneous spoken language. Clearly, the regimens of ketamine used in the experiment did not induce changes in verbal productivity so conspicuous as to be unmistakably detected by a human audience. It is therefore imperative that we use the computer for an accurate and objective assessment of the data.

I chose to analyze the data from ten of the volunteers, for whom I have all the three recordings. The speech samples from the other ten volunteers were left untouched to be used later in an unprejudiced test of KSAS. I generated corresponding .f0, .frm and transcription files for each of the thirty speech samples from the selected volunteers before inputting them to KSAS. I then obtained some statistics from KSAS for further study.

3.2 Statistical Results

For most of the measures output by KSAS, I did a paired one-tail Student’s *t*-test between the set of high-dose values and that of placebo values, and between the set of low-dose values and the placebo values. The results of the *t*-tests show that values of some measures were significantly altered with the administration of ketamine (the *p*-values reported below are from the tests between Regimen A (i.e., high dose) and placebo). These measures include:

1) Percentage of verb tokens: Ketamine significantly reduces the frequency of verb tokens ($p<0.003$);

2) Occurrences of fillers like *um*, *uh* per one hundred words: ketamine significantly increases the use of fillers ($p<0.011$);

3) F1 interquartile range (iqr) as percentage of the total F1 range: Ketamine significantly increases the F1 iqr as a percentage of the total F1 range ($p<0.013$);

4) Percentage of noun types: Ketamine significantly increases the percentage of noun types ($p<0.014$);

5) Percentage of noun tokens: Ketamine significantly increases the percentage of noun tokens ($p<0.037$);

6) F1 standard deviation: Ketamine significantly increases the standard deviation of F1 ($p<0.057$).

Table 3.1 reports the detailed experimental results of the above six significant measures.

Below are some other measures that are affected by ketamine though not as significantly; Table 3.2 reports the experimental results for these measures.

1) Number of occurrences of *I* per 100 words: Ketamine increases the frequency of the pronoun *I* ($p<0.11$);

2) Frequency of positive assertions: Ketamine increases the frequency of *yes*, *yeah* and *yep* ($p<0.15$);

3) F0 standard deviation: Ketamine increases the standard deviation of F0 ($p<0.13$);

4) Number of questions: Ketamine reduces the number of questions ($p<0.14$);

5) F3 range: Ketamine reduces the range of F3 ($p<0.13$).

Along with the increase in the standard deviation of F0 and F1 mentioned above, an insignificant decrease in the kurtosis of F0 and F1 after ketamine has been observed. Table 3.3 shows the changes in these two values of F0 and F1.

Regimen	Volunteer ID (with Regimen Set ID)										*p*-value (Ketamine paired with Placebo)
	1 (1)	3 (1)	4 (1)	5 (1)	7 (1)	9 (1)	10 (1)	11 (1)	14 (2)	15 (2)	
Percentage of verb tokens											
A	25.78	30.67	32.08	34.72	30.45	31.7	29.44	33	31.82	32.73	0.003
B	28.51	30.51	32.1	34.72	30.54	33.76	29.57	28.63	29.2	28.81	0.0001
C (Placebo)	33.16	32.74	35.29	34.22	32.97	36.49	33.1	33.23	32.31	34.87	
Occurrences of fillers like *um*, *uh* per one hundred words											
A	7.63	3.11	1.73	4.56	3.7	3.77	5.22	7.26	6.82	6.07	0.010
B	4.82	5.88	0.53	4.56	3.59	2.06	4.13	5.65	5.32	6.58	0.042
C (Placebo)	3.65	1.77	1.61	3.17	4.96	0.61	3.45	3.86	3.08	6.91	
F1 interquartile range (iqr) as percentage of the total F1 range											
A	13.61	26.57	20.88	17.46	20.72	17.81	12.34	14.87	10.71	14.15	0.013
B	21.23	24.18	15.78	17.74	22.04	21.55	8.56	21.08	8.26	13.16	0.016
C (Placebo)	9.24	14.01	13.32	15.56	21.01	12.44	8.59	15.39	9.07	15.15	
Percentage of types that are nouns											
A	25	20.31	21.01	17.45	22.55	26.15	20.11	16.86	22.08	20.54	0.014
B	22.22	23.21	19.29	17.45	20	21.94	23.49	19.23	20	26.77	0.008
C (Placebo)	18.59	20.51	19.59	16.06	14.48	18.42	14.78	18.92	21.82	19.12	
Percentage of tokens that are nouns											
A	25.78	26.67	19.08	18.98	23.46	20.75	19.83	16.01	21.21	23.02	0.037
B	23.69	24.58	20.42	18.98	20.96	22.16	20	21.77	24.34	24.69	0.003
C (Placebo)	16.97	25.66	19.79	14.81	19.51	15.32	18.28	18.69	22.56	21.05	
F1 standard deviation											
A	168.86	167.4	147.4	222.67	178.22	164.36	137.62	115.89	90.42	168.13	0.057
B	186.91	164.46	156.03	158.03	186.46	167.72	119.14	166.49	123.15	119.42	0.013
C (Placebo)	164.67	159.24	126.19	147.59	184.3	118.89	134.89	124.68	113.43	111.32	

Table 3.1 Most significant experimental results

Regimen	Volunteer ID (with Regimen Set ID)										p-value (Ketamine paired with Placebo)
	1 (1)	3 (1)	4 (1)	5 (1)	7 (1)	9 (1)	10 (1)	11 (1)	14 (2)	15 (2)	
Number of occurrences of *I* per 100 words											
A	0.7	1.78	0.87	1.37	2.88	0.19	2.71	0.66	0	1.8	0.110
B	0.4	2.12	1.59	1.37	2.7	0	1.96	0	0.44	0.41	0.371
C (Placebo)	0.79	1.33	0.54	1.06	2.47	0.2	0.69	0.89	0.51	1.64	
Frequency of positive assertions											
A	0.7	0.44	0.58	0.46	1.23	0	0.63	0.17	0	0.72	0.146
B	0.4	0	1.06	0.46	1.8	0	0.43	0	0	0.41	0.253
C (Placebo)	0	1.7	0	0	0	0.2	0	0	0	0.66	
F0 standard deviation											
A	19.95	27.96	21.11	28.56	29	15.82	25.73	16.03	22.47	26.93	0.132
B	32.12	22.95	16.7	14.3	27.39	22.16	16.54	44.58	16.21	29.25	0.209
C (Placebo)	21.24	21.88	33.55	17.08	29.29	13.95	17.27	13.63	18.28	22.96	
Number of questions											
A	2	1	0	0	1	0	0	0	0	0	0.139
B	0	0	0	0	0	0	1	0	0	0	0.056
C (Placebo)	1	2	0	0	2	0	0	0	0	2	
F3 range											
A	2211.8	2408.4	2321.1	2820.9	2953.8	2692.8	2791.5	2595.8	2388.1	2312.3	0.126
B	2077.5	2109.5	2207.7	2869.3	3122.7	2398.9	2968.5	2813.8	2566.8	2215.9	0.037
C (Placebo)	2083.5	2446.9	2255	3138.8	2992	2392	3096.2	2904.5	2537	2446.7	

Table 3.2 Some results with lower significance

Regimen	Volunteer ID (with Regimen Set ID)										*p*-value (Ketamine paired with Placebo)
	1 (1)	3 (1)	4 (1)	5 (1)	7 (1)	9 (1)	10 (1)	11 (1)	14 (2)	15 (2)	
F0 standard deviation											
A	19.95	27.96	21.11	28.56	29	15.82	25.73	16.03	22.47	26.93	0.132
B	32.12	22.95	16.7	14.3	27.39	22.16	16.54	44.58	16.21	29.25	0.209
C (Placebo)	21.24	21.88	33.55	17.08	29.29	13.95	17.27	13.63	18.28	22.96	
F0 Kurtosis											
A	41.4	25.31	29.63	29.38	11.18	106.18	15.16	71.82	10.41	26.37	0.201
B	22.81	19.81	26.57	19.68	26.52	52.79	35.66	15.77	9.25	32.6	0.044
C (Placebo)	31.09	25.29	22.41	41.75	27.2	116.02	25.14	56.45	12.23	39.97	
F1 standard deviation											
A	168.86	167.4	147.4	222.67	178.22	164.36	137.62	115.89	90.42	168.13	0.057
B	186.91	164.46	156.03	158.03	186.46	167.72	119.14	166.49	123.15	119.42	0.013
C (Placebo)	164.67	159.24	126.19	147.59	184.3	118.89	134.89	124.68	113.43	111.32	
F1 Kurtosis											
A	5.95	2.7	3.2	6.77	4.53	4.08	12.91	6.55	11.61	9.44	0.259
B	3.72	2.74	4.29	4.66	4.26	2.81	17.31	5.28	21.06	10.71	0.410
C (Placebo)	6.84	3.98	5.6	5.46	3.66	4.91	18.69	8.13	12.74	3.89	

Table 3.3 Standard deviation and kurtosis of F0 and F1

3.3 Discussion

In this section, I discuss the distinctive properties of ketamine-influenced speech found in the last section.

3.3.1 Verbs vs. Nouns

Compare the utterances in examples (1) and (2) below which are taken from the transcriptions of the speech samples by two of the volunteers. In both examples, utterance (a) was produced after ketamine and utterance (b) after placebo.

(1) (a) "There's a table. The – the other female figure is older – could either be a, a maid, servant or mother." (Volunteer A – ketamine-influenced speech)

(b) "There's um, looks like this lady in the foreground who's, looks like she's probably up a tree. Looking down on another lady … And the woman looks as though she's running. Perhaps searching. Searching for her. Her eyes look quite piercing. As though she's looking out for something." (Volunteer A – normal speech)

(2) (a) "And a woman looking over a bridge. With, um, docks and housing in the background. With people and the docklands. And unloading a boat." (Volunteer B – ketamine-influenced speech)

(b) "Ok, um, appear to be a sort of beach scene. And there's two women in the picture. Um looks as if there's water in the background. The woman in the bottom left is slightly, slightly more distant. The woman in the foreground appears to be halfway up a tree or tree trunk." (Volunteer B – normal speech)

Utterance (a) in both examples makes a striking impression of the speaker naming one object after another, with no further intention of describing their actions. The ketamine-influenced speech is stacked with sentence fragments made up of noun phrases, which are not extended with verb phrases. On the other hand, in normal speech such as (b), sentences are frequently started with a verb, with the subjects (expletive or non-expletive) omitted. The speakers not only attempt to describe the actions depicted in the picture, but tend to modify and revise their verbs. Moreover, repairs are more frequently made inside a noun phrase in ketamine-influenced speech, while they are often made as a verb is being uttered in normal speech, as shown in example (1).

The increase of the percentage of verbs and the decrease of the percentage of nouns in ketamine-influenced speech is fundamentally caused by some brain impairment, reflected in the difficulty in retrieving the more complicated lexical items. Such impairment is exaggerated with high doses of ketamine, as described by a heavy ketamine-taker as "trouble searching for the right word, and trouble putting sentences together" (Rocher 2001).

Verbs are intrinsically more complex than nouns. Verbs are semantically related to their arguments and syntactically retrieved with their argument structures, whereas typically nouns simply refer to single entities and serve as arguments. Verbs have less imageability, since they refer to actions and events that usually change over time rather than stay static (D'Amico et al. 2002). The greater complexity of verbs is supported by evidence from language acquisition, since all children acquire nouns earlier than verbs. The fact that verbs are more complex than nouns implies that any brain pathology should be more likely to impair verbs than nouns

(Silveri et al. 2003). In fact, recent research has shown proven relationship between verb production impairments and mental disorders such as dementia (Silveri et al. 2003, Bak et al. 2001), Alzheimer's disease (Kim and Thompson 2004), and aphasia (Berndt et al. 2002).

I conclude that subanesthetic doses of ketamine cause minor impairment that subtly hinders the retrieval of verbs, while leaving noun production unaffected. This implies that ketamine may affect certain parts of the nervous system compromised by the mental disorders whose symptoms include verb production disorder.

3.3.2 The First Formant and the Fundamental Frequency

One thing that ketamine-takers often report when on ketamine is having problem "making the sounds" (Rocher 2001), one typical example being "the words 'loads of love' come out as 'roads of ruv'" (StoneHappyMonday 2001). This indicates that while on ketamine, the central nervous system has difficulty controlling the articulatory organs. Since formant frequencies convey place and manner of articulation, deviations from the formant distributions of normal speech serve as a helpful indicator of articulatory difficulty. For example, unusual distribution of F1 reflects abnormal vowel height resulting from articulatory irregularity.

In this study, ketamine significantly altered the standard deviation of F1 and its interquartile range expressed as a percentage of the total range. An insignificant decrease in the kurtosis of F1 accompanied the increase of the standard deviation after ketamine. However, there was no significant change in the range and mean of F1. These results show that the first formant of normal speech is more concentrated and peaked at the center of the range while that of ketamine-influenced speech is more spread out. A plausible explanation of this contrast is that the brain partially loses control of the articulators (especially those regulating vowel height) and has trouble finding the "anchor" of normal articulation.

A few of the previous studies have examined the relationship between spectral changes and mental diseases such as schizophrenia (Tolkmitt et al. 1982), Down's Syndrome (Moran 1986; Pentz 1987) and speech disorders (Kent 1979). In those studies, formant changes are examined only with regard to a few specific vowels and consonants. In addition, changes in the second formant are more frequently reported in the literature (e.g., Kato 1989; Subramanian 2003). My study differs from the previous studies in that I did not isolate certain sounds for examination and I found that the first formant experienced more prominent changes than the other formants. This result coincides to a certain degree with the finding of Haley, Ohde and Wertz (2001), who claimed that vowel substitutions that some aphasic and apraxic speakers displayed affected exclusively vowel height rather than vowel frontness (as partially represented by F2).

I had expected to detect "affective flattening," and specifically "lack of vocal inflections" (i.e., monotonous speech with flattened intonation), through decreased variance in the distribution of the fundamental frequencies. On the contrary, like the first formant, the fundamental frequencies in ketamine-influenced speech were found to generally exhibit an increased standard deviation and a decreased kurtosis in my project. My speculation is that the brain on ketamine is also having trouble anchoring the pitch, and that the ketamine doses in this project might not have been high enough to significantly reduce vocal inflections.

3.3.3 Filled and Unfilled Pauses

Alogia is an important sign of negative-symptom schizophrenia, which is characterized by poverty of speech and increased latency of responses. Ketamine has been reported to simulate schizophrenia-like negative symptoms in healthy volunteers (e.g., Adler et al. 1998; Lahti et al. 1995, 2001) and I have shown that the frequency of filled pauses significantly increases with ketamine. Such increase may be either due to the difficulty in lexical retrieval mentioned earlier in 3.3.1, or due to poverty of content of speech, or due to drug-induced emotional withdrawal.

Contrary to my expectation, however, my measure for unfilled pauses, the percentage of time with zero F0 (i.e., no voicing), failed to produce a very significant result, although studies have shown negative symptoms to be highly correlated with the frequency and length of unfilled pauses (Alpert, Pouget and Silva 1995; Stassen et al. 1995). This is because the values of this measure are severely compromised in this project by the interviewers in three ways: first, some interviewers conversed with the volunteers while others did not talk at all during the picture descriptions; second, some interviewers prompted for more description after the volunteer finished talking about one picture (which was obviously difficult for the volunteer) while others proceeded right to the next picture; last and most significantly, the pauses that the interviewers left between the pictures varied vastly. Hence, I still believe that the percentage of time with zero F0 or related F0 measures such as the average length of pauses with zero F0 could be useful on other occasions.

3.3.4 Some Pragmatic Measures

I found a tendency to increase the use of the pronoun *I* in ketamine-influenced speech. This is apparently related to the “self-reference” disorder used in the TLC scale to rate thought disorder (Andreasen 1979a). Another pragmatic measure that has been proved useful is the number of questions in the speech. Questions are, in a way, an acknowledgment of the existence of the interlocutor and a way to establish a relationship. Ketamine-influenced speech is more self-centered and contains fewer questions.

Although there is no obvious explanation available for these results, they reveal to a certain extent impairment of interpersonal communicative skills or social inference that is often referred to as “theory of mind” (Premack and Woodruff 1978; Corcoran, Mercer and Frith 1995, Frith 2004).

CHAPTER FOUR

THE SCORING AND IDENTIFICATION SYSTEM

The statistical analyses in the previous chapter reveal some distinctive properties of ketamine-influenced speech. In this chapter, I develop a simple algorithm for KSAS, which combines these properties into a weighted score, and use the score to distinguish ketamine-influenced from normal speech.

4.1 A Simple Algorithm to Score and Identify

On clicking the “conclusion” button, KSAS will present in its conclusion text box its suggestion as to which one of the input speech samples is ketamine-influenced and which one is normal. This is achieved by running a simple algorithm that scores the speech samples and does the identification based on the characteristics of ketamine-influenced speech discussed in 3.2.

I chose to include ten of the distinctive properties in the scoring system:

1) Lower percentage of verb tokens;
2) Higher percentage of noun types;
3) Higher frequency of fillers like *um*, *uh*;
4) Higher F1 interquartile range as percentage of the total F1 range;

5) Larger F1 standard deviation;

6) Larger F0 standard deviation;

7) Smaller number of questions;

8) More occurrences of *I* per 100 words;

9) Higher frequency of positive assertions;

10) Larger F3 range.

"Lower percentage of noun tokens" was excluded from the list, because it is highly correlated with "lower percentage of noun types" and would cause multi-collinearity in the scoring system.

The basic strategy used to score the speech samples and identify ketamine-influenced speech is:

1) Compare the two input speech samples; for the presence of each of the distinctive properties found in a speech sample, add to the total score of that sample;

2) Compare the total scores of the two input speech samples, and label the sample with the higher score as "ketamine-influenced."

The ten properties, however, have different levels of significance as shown in 3.2 by different *p*-values; it would, therefore, be inappropriate simply to assign 1 point for each of the properties. In order to obtain indicative scores with a reasonable weight assigned to each property, I did two things.

First, since the first five properties are supported by much more significant test results than the latter five, I separated the ten properties into two groups: the more significant and the less significant group. The second group will be used only if there is a tie in the scores from using the first group. Four properties from the second group (i.e., F0 standard deviation, number of questions, frequency of *I*, and frequency of positive assertion) are almost always able to break the tie; in the very rare case where there is another tie (which did not happen at all in my experiment), the last property “F3 range” will be used to make the identification.

Second, the score assigned to each property was roughly weighted according to its p-values both from the paired t-test on high-dose ketamine (Regimen A) and placebo samples, and the paired t-test on low-dose ketamine (Regimen B) and placebo samples. Briefly, the following shows how the weights are assigned to the properties:

1) with the smallest *p*-values ($p<0.003$), “lower percentage of verb tokens” will score 1.5 points;
2) with their *p*-values around 0.01, “higher percentage of noun types,” “higher frequency of fillers like ‘um’, ‘uh’,” and “higher F1 interquartile range as percentage of the total F1 range” will each score 1 point;
3) with its *p*-values around 0.05, “larger F1 standard deviation” will score 0.5 points;
4) the remaining five less significant properties will each score 0.2 points.

Assuming that the values of the necessary measures have already been computed for the input speech samples, the scoring and identification algorithm is basically a series of number comparisons and additions, and thus extremely efficient. Appendix B is the actual implementation of the algorithm.

4.2 Results

The scoring and identification algorithm is simple, yet it achieves high accuracy in identifying ketamine-influenced speech. I tested the algorithm on all the 37 contrastive pairs of speech samples at hand, each consisting of one normal and one ketamine-influenced sample produced by a single volunteer. The whole process, from data extraction to conclusion, was automated, and the identification algorithm successfully recognized the ketamine-influenced sample from 36 of the speech pairs, which means an accuracy of 97%. Noticeably, 83% of the correct identifications were achieved by using only the five most significant properties. Table 4.1 is a detailed report of the test results.

These results are significant especially in two aspects. First, I have thereby proved that subanesthetic doses of ketamine do induce deviations from normal speech, if not speech impairments. These differences occur at all levels of language: phonetic, semantic, syntactic and pragmatic. Second, differences between ketamine-influenced and normal speech have been proved machine detectable. While human observers may have certain advantage over machines in perceiving emotional changes in a face-to-face interview, some subtle changes in the speech such as a slight decrease in the percentage of verbs are more liable to be detected by machines. The contrast between humans and machines in my project boils down to the contrast between subjectivity and objectivity and the contrast between fuzziness and precision. I believe that if objectivity and precision are used in complement to expert judgment, our conclusions shall be closer to fact in any psychiatric research.

Vol ID	Reg set ID	Paired regimens	No. of props used	Paired scores	Result
1	1	A-C	5	5-0	Success
1	1	B-C	5	5-0	Success
2	1	B-C	9	2.7-2.5	Success
3	1	A-C	5	4-1	Success
3	1	B-C	5	5-0	Success
4	1	A-C	5	5-0	Success
4	1	B-C	5	3-2	Success
5	1	A-C	5	3.5-1.5	Success
5	1	B-C	5	5-0	Success
6	1	A-C	5	4-1	Success
6	1	B-C	5	3.5-1.5	Success
7	1	A-C	9	3.1-2.7	Success
7	1	B-C	5	4-1	Success
9	1	A-C	5	5-0	Success
9	1	B-C	5	5-0	Success
10	1	A-C	5	5-0	Success
10	1	B-C	5	3.5-1.5	Success
11	1	A-C	9	2.9-2.7	Success
11	1	B-C	5	5-0	Success
12	1	A-C	5	5-0	Success
12	1	B-C	5	4-1	Success
13	2	B-C	5	3-2	Success
14	2	A-C	5	4.5-0.5	Success
14	2	B-C	5	3-2	Success
15	2	A-C	5	3-2	Success
15	2	B-C	5	3-2	Success
18	2	B-C	5	4-1	Success
20	2	A-C	5	5-0	Success
20	2	B-C	5	5-0	Success
21	2	A-C	9	2.9-2.5	Success
21	2	B-C	9	2.9-2.5	Success
22	2	A-C	5	**1.5-3.5**	**Failure**
22	2	B-C	5	4-1	Success
23	2	A-C	5	5-0	Success
23	2	B-C	5	4-1	Success
24	2	A-C	9	2.9-2.7	Success
24	2	B-C	5	3.5-1.5	Success

Table 4.1 KSAS test results

CHAPTER FIVE

CONCLUSION

Ketamine, as an NMDA antagonist, has been known to induce brief schizophrenia-like symptoms in healthy volunteers. Since many of these symptoms concern language, this study is devoted to finding distinctive properties of ketamine-influenced speech, in the hope that such findings may shed light on the characteristics of schizophrenic language and thus help in the diagnosis of schizophrenia.

This study shows that ketamine-influenced speech differs significantly from normal speech in at least five aspects and these have successfully been used to identify ketamine-influenced speech automatically:

1) ketamine significantly reduces the percentage of verbs in the speech;
2) ketamine significantly increases the percentage of nouns;
3) ketamine significantly increases the use of fillers like *um*, *uh*;
4) ketamine significantly increases the F1 interquartile range as a percentage of the total F1 range;
5) ketamine significantly increases the standard deviation of F1.

An obvious step to take in the future is to test the properties found to be significant for ketamine-influenced speech on data from schizophrenic patients. Once these or some of these properties are proven to be shared by schizophrenic speech, they can be adopted in diagnostic procedures for schizophrenia. Changes towards normal speech should provide valuable cues for the improvement of a patient's condition.

As a by-product of this study, the Ketamine Speech Analysis Software (KSAS) has been implemented in Java. It extracts basic statistics at phonetic, semantic, syntactic and pragmatic level from the input speech samples, which can be used in further study. In particular, it includes a scoring and identification algorithm that distinguishes ketamine-influenced from normal speech with a high accuracy. Following the same steps that I took to develop the algorithm, KSAS can be easily adapted to identify typical speech induced by other pathologies. All that needs to be done is to find the statistically significant properties and rewrite the scoring system. Obviously, the usefulness of the software is not restricted merely to the study of ketamine-influenced speech.

One of the improvements planned for the software is the implementation of a built-in syntactic parser, which will make it possible to rate sentence complexity. Since, as I have mentioned in 2.3, mental diseases often reduce syntactic complexity, it will be a useful measure to have. Another move to make may be to revise the scoring system to make it more sophisticated and more error-proof.

BIBLIOGRAPHY

Adler, Caleb M., Terry E. Goldberg, Anil K. Malhotra, David Pickar and Alan Breier (1998) "Effects of Ketamine on Thought Disorder, Working Memory and Semantic Memory in Healthy Volunteers," *Biological Psychiatry*, 43, 811-816.

Allen, H. A., P. F. Liddle and C. D. Frith (1993) "Negative Features, Retrieval Processes and Verbal Fluency in Schizophrenia," *British Journal of Psychiatry*, 163, 769–775.

Alpert, Murray, Enrique R. Pouget and Raul Silva (1995) "Cues to the Assessment of Affects and Moods: Speech Fluency and Pausing," *Psychopharmacology Bulletin*, 31(2), 421-424.

American Psychiatric Association (2000) *Handbook of Psychiatric Measures*. Washington, D.C.: American Psychiatric Association.

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4^{th} ed.). Washington, D.C.: American Psychiatric Association.

Andreasen, Nancy C. (1984) *Scale for the Assesssment of Positive Symptoms (SAPS)*. Iowa City, IA: University of Iowa, 1984.

Andreasen, Nancy C. (1982) "Negative Symptoms in Schizophrenia: Definition and Reliability," *Archives of General Psychiatry*, 39, 784-788.

Andreasen, Nancy C. (1979a) "Thought, Language, and Communication Disorders," *Archives of General Psychiatry*, 36, 1315-1321.

Andreasen, Nancy C. (1979b) "Affective Flattening and the Criteria for Schizophrenia," *American Journal of Psychiatry*, 136(7), 944-947.

Bak, Thomas H., Dominic G. O'Donovan, John H. Xuereb, Simon Boniface and John R. Hodges (2001) "Selective Impairment of Verb Processing Associated with Pathological Changes in Brodmann Areas 44 and 45 in the Motor Neurone Disease-Dementia-Aphasia Syndrome," *Brain*, 124, 103-120.

Berndt, Rita Sloa, Martha W. Burton, Anne N. Haendiges and Charlotte C. Mitchum (2002) "Production of Nouns and Verbs in Aphasia: Effects of Elicitation Context," *Aphasiology*, 16(1/2), 83-106.

Center for Speech Technology (2004) *WaveSurfer* (v1.6.2), Stockholm, Sweden, http://www.speech.kth.se/wavesurfer/.

Chance, S. A., S. L. Collinson, A. C. James and T. J. Crow (2002) "Abnormality of the Semantic Network in Adolescent Subjects with Schizophrenia," http://www.psychiatry.ox.ac.uk/powic/posters/stevenchancewws2002.htm.

Corcoran, Rhiannon, Gavin Mercer and Christopher D. Frith (1995) "Schizophrenia, Symptomatology and Social Inference: Investigating 'Theory of Mind' in People with Schizophrenia," *Schizophrenia Research*, 17, 5-13.

D'Amico, Simona, Simone Bentrovato, Manual Gasparini, Daniela Costabile and Elizabeth Bates (2002) "Timed Picture Naming in Italian-Speaking Children and Adults: Differences Between Nouns and Verbs," http://crl.ucsd.edu/~bates/papers/pdf/iascl-srcld/srcld.action.object.short.pdf.

Didion, John et al. (2003) *JWNL* (Software, v1.3), Stanford University, http://sourceforge.net/projects/jwordnet.

Frith, Christopher D. (2004) "Schizophrenia and Theory of Mind," *Psychological Medicine*, 34(3), 385-389.

Haley, Katarina L., Ralph N. Ohde and Robert T. Wertz (2001) "Vowel Quality in Aphasia and Apraxia of Speech: Phonetic Transcription and Formant Analyses," *Aphasiology*, 15(12), 1107-1123.

Kato, Yasuyoshi and Tomoyoshi Yoshino (1989) "Second Formant Transitions in the Speech of Persons with Profound Hearing Impairment," *Japanese Journal of Special Education*, 27(2), 75-81.

Kent, Ray D. (1979) "Isovowel Lines for the Evaluation of Vowel Formant Structure in Speech Disorders," *Journal of Speech and Hearing Disorders*, XLIV, 513-521.

Kim, Mikyong and Cynthia K. Thompson (2004) "Verb Deficits in Alzheimer's Disease and Agrammatism: Implications for Lexical Organization," *Brain and Language*, 88(1), 1-20.

Jansen, Karl (2002) "Ketamine and the People," http://www.drugtext.org/library/articles/jansen02.html.

Kåre, Sjölander (2002) "Recent Developments Regarding the WaveSurfer Speech Tool," Presented at the 2002 Annual Swedish Phonetics Meeting, Stockholm, Sweden, May 29–31, http://www.speech.kth.se/qpsr/tmh/2002/02-44-053-056.pdf.

Ladefoged, Peter (2001) *A Course in Phonetics* (4th ed.). Orlando, Florida: Harcourt, Inc.

Lahti, Adreinne C., Bettylou Koffel, David LaPorte and Carol A. Tamminga (1995) “Subanesthetic Doses of Ketamine Stimulate Psychosis in Schizophrenia,” *Neuropsychopharmacology*, 13(1), 9-19.

Lahti, Adrienne C., Martin A. Weiler, Tamara Michaelidis, Arti Parwani and Carol A. Tamminga (2001) “Effects of Ketamine in Normal and Schizophrenic Volunteers ,” *Neuropsychopharmacology*, 25(4), 455-467.

Malhotra, Anil K., Debra A. Pinals, Herbert Weingartner, Karen Sirocco, C. David Missar, David Pickar and Alan Breier (1996) “NMDA Receptor Function and Human Cognition: The Effects of Ketamine in Healthy Volunteers,” *Neuropsychopharmacology*, 14(5), 301-307.

Marshall, Jane (2003) “Noun-Verb Dissociations – Evidence from Acquisition and Developmental and Acquired Impairments,” *Jounnal of Neurolinguistics*, 16(2/3), 67-84.

McCandless, S. S. (1974) “An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-22(2), 135-141.

Meyer, Peter (1989) “Report on a Ketamine Experience,” http://www.serendipity.li/dmt/ket89.html.

Miller, George A. et al. (2003) *WordNet* (Software, v2.0), Cognitive Science Laboratory, Princeton University, http://www.cogsci.princeton.edu/~wn/index.shtml.

Moran, Michael J. (1986) “Identification of Down’s Syndrome Adults from Prolonged Vowel Samples,” *Journal of Communication Disorders*, 19, 387-394.

Morice, Rodney and J. C. L. Ingram (1982) “Language Analysis in Schizophrenia: Diagnostic Implications,” *Australian and New Zealand Journal of Psychiatry*, 16, 11–21.

Morice, Rodney and Don McNicol (1985) “The Comprehension and Production of Complex Syntax in Schizophrenia,” *Cortex*, 21, 567-580.

Morice, Rodney and Don McNicol (1986) “Language Changes in Schizophrenia: A Limited Replication,” *Schizophrenia Bulletin*, 12, 239-251.

Murray, H. A. (1971). *Thematic Apperception Test: Manual.* Cambridge, MA: Harvard University Press. (Original work published 1943)

National Institute of Mental Health (1999) “Schizophrenia,” http://www.nimh.nih.gov/publicat/schizoph.htm.

Overall, John E. and Donald R. Gorham (1962) “The Brief Psychiatric Rating Scale,” *Psychological Reports*, 10, 799-812.

Pentz, Arthur L., Jr. (1987) “Formant Amplitude of Children with Down Syndrome,” *American Journal of Mental Deficiency*, 92(2), 230-233.

Premack, D. and G.. Woodruff (1978) “Does the Chimpanzee Have a Theory of Mind?” *Behavioral and Brain Sciences*, 1(4), 515-526.

Rocher (2001) “Ketamine and Speaking Problems” (internet post), http://www.bluelight.nu/vb/showthread.php?threadid=72091.

Secrest, Bruce G. and George R. Doddington (1983) “An Integrated Pitch Tracking Algorithm for Speech Systems,” *Proceedings of ICASSP’83*, IEEE Signal Processing Society, 1352-1355.

Silveri, M. C., B. L. Salvigni, A. Cappa, C. Della Vedova and M. Puopolo (2003) "Impairment of Verb Processing in Frontal Variant-Frontotemporal Dementia: A Dysexecutive Symptom," *Dementia and Geriatric Cognitive Disorders*, 16(4), 296-300.

Spitzer, R. L., J. Endicott and E. Robins (1975) *Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders.* New York, NY: New York State Psychiatric Institute.

Stassen, H. H., M. Albers, J. Puschel, C. H. Scharfetter, M. Tewesmeier and B. Woggon (1995) "Speaking Behavior and Voice Sound Characteristics Associated with Negative Schizophrenia," *Journal of Psychiatric Research*, 29(4), 277-296.

Stein, Johanna (1993) "Vocal Alterations in Schizophrenic Speech," *The Journal of Nervous and Mental Disease*, 180(1), 59-62.

StoneHappyMonday (2001) "Ketamine and Speaking Problems" (internet post), http://www.bluelight.nu/vb/showthread.php?threadid=72091.

Subramanian, Anu (2003) "Second Formant Transitions in Fluent Speech of Persistent and Recovered Preschool Children Who Stutter," *Journal of Communication Disorders*, 36(1), 59-75.

Talkin, David (1995) "A Robust Algorithm for Pitch Tracking (RAPT)," in Kleijn, W. B. and K. K. Paliwal (eds.), *Speech Coding and Synthesis*, New York, NY: Elsevier, 495-518.

Tolkmitt, F., H. Helfrich, R. Standke and K. R. Scherer (1982) "Vocal Indicators of Psychiatric Treatment Effects in Depressives and Schizophrenics," *Journal of Communication Disorders*, 15, 209-222.

Van Riezen, Henk and Maria C. Vrijmoed-de Vries (2000) *Rating Scales for Psychoneuropharmacology* (CD-ROM edition), Elsevier.

White, Paul F., Walter L. Way and Anthony J. Trevor (1982) "Ketamine – Its Pharmacology and Therapeutic Uses," *Anaesthesiology*, 56, 119-136.

World Health Organization (1992) *The ICD-10 Classification of Mental and Behavioural Disorders*, Geneva.

APPENDIX A

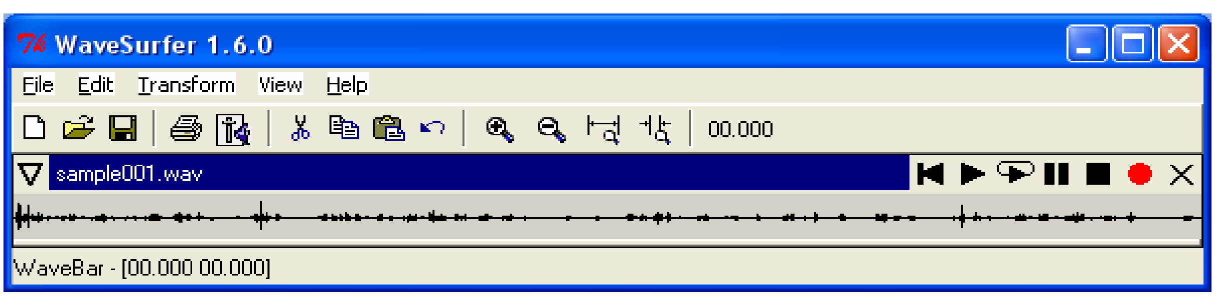
HOW TO USE KSAS

In this appendix, I use an example to walk the reader through the whole process of using KSAS.

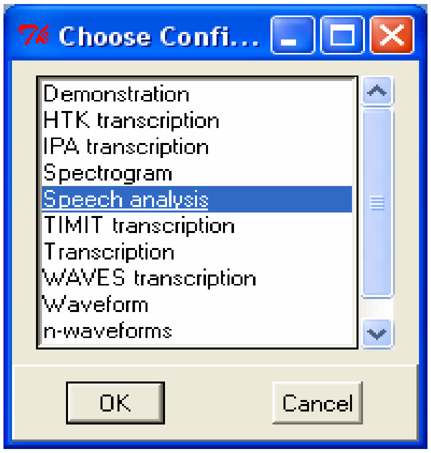
Say we have in folder “Volunteer-A” two sound files sample001.wav and sample002.wav, which are recordings of one healthy volunteer A. One of the files was recorded after ketamine, and the other was not. Remember that if the sound files contain the voice of an interviewer, they need editing so that the files contain only the voice of the volunteer. Follow these steps to use KSAS:

1. Create two subfolders “sample001” and “sample002” inside the folder “Volunteer-A.” Move sample001.wav and sample002.wav to their corresponding subfolders.
2. Transcribe sample001.wav to sample001.txt. Do not include comments. Save sample001.txt in the folder “sample001.”

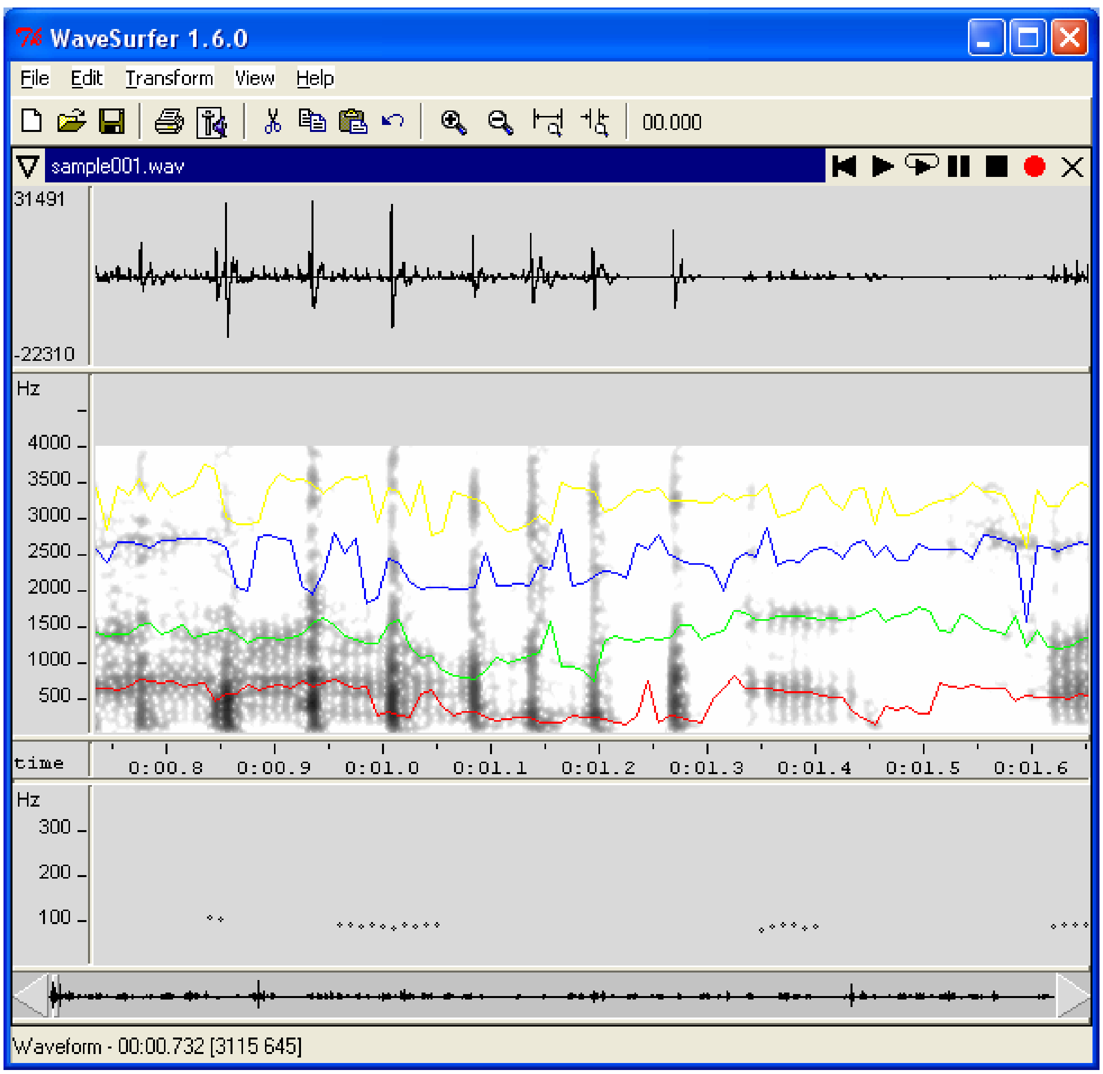
Now, open sample001.wav using WaveSurfer.



When prompted, choose "Speech analysis" as the configuration and click OK.



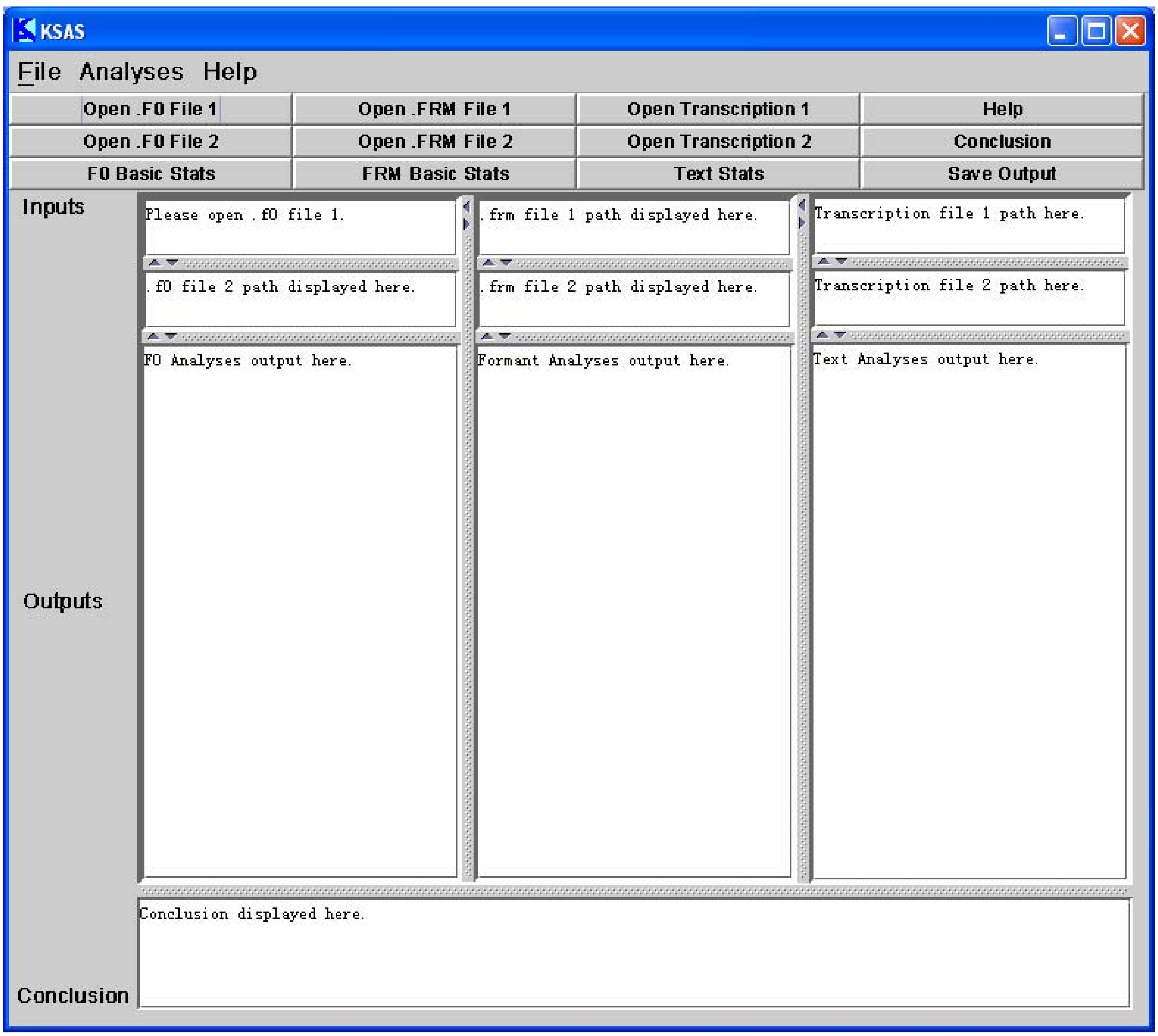
Wait for WaveSurfer to complete its computation.



Right click on the spectrogram pane and choose “Save Data File ...” from the menu. Save the file “sample001.frm” to the folder “sample001.” Right click on the pitch contour pane and choose “Save Data File …” from the menu. Save the file “sample001.f0” to the folder “sample001.”

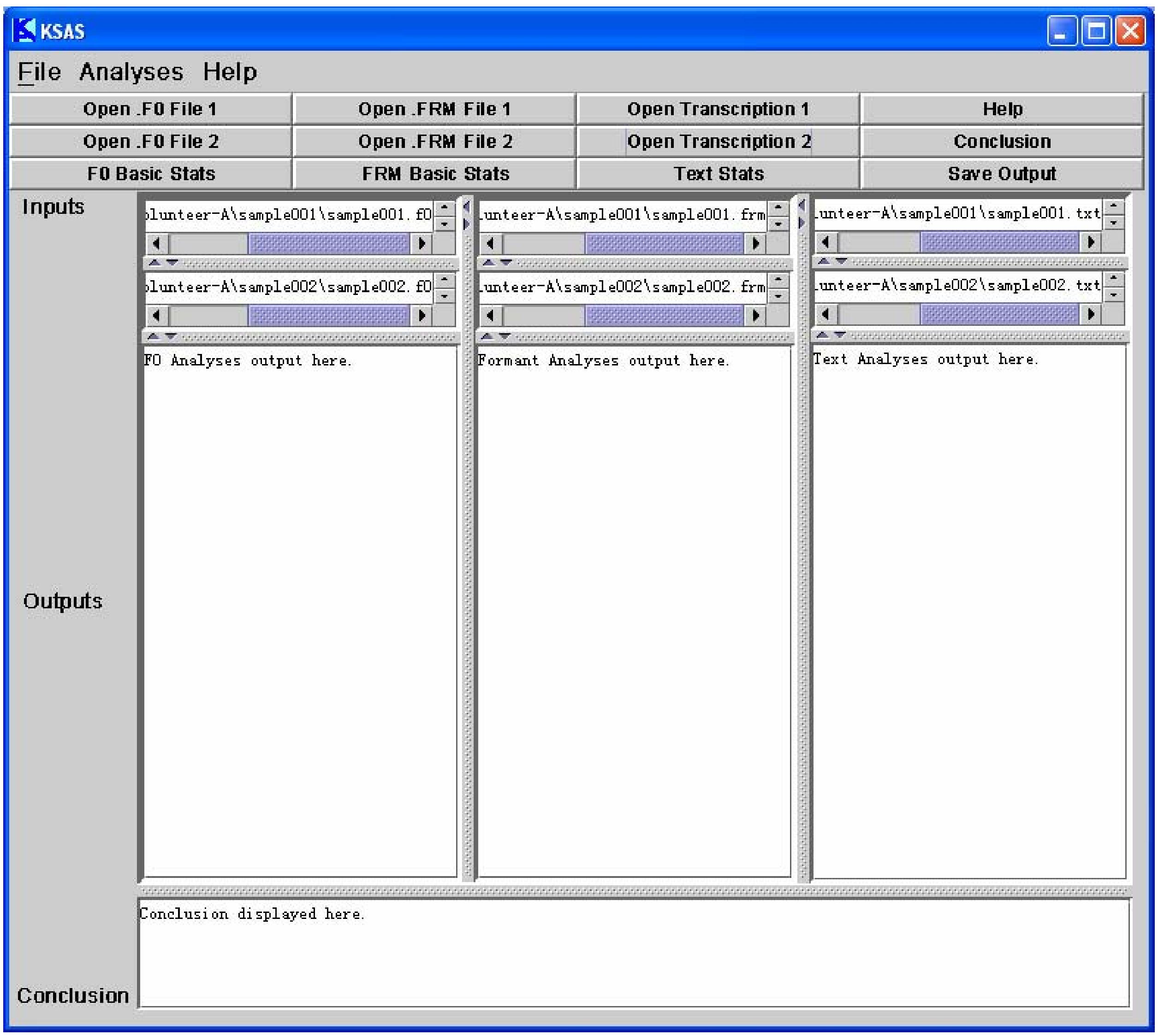
3. Repeat step 2 for sample002.wav in the folder “sample002.” Now, we should have two folders each containing four files: a .wav, a .f0, a .frm and a .txt file.

4. Open KSAS.

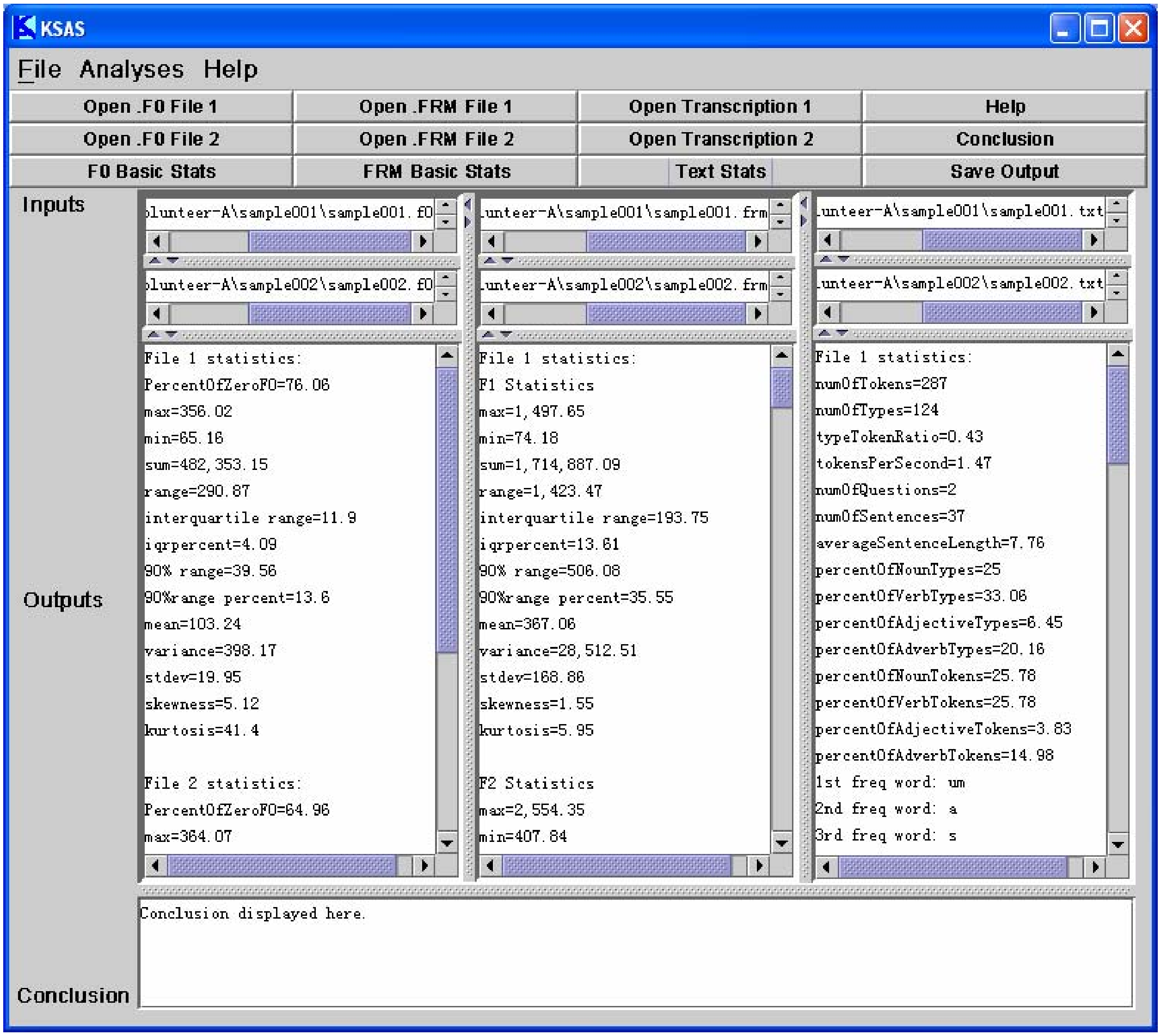


5. Click the "Open .F0 File 1" button and open "sample001.f0." Click the "Open .FRM File 1" button and open "sample001.frm." Click the "Open Transcription 1" button and open "sample001.txt." With each file opened, the corresponding input window will display the file name and its path.

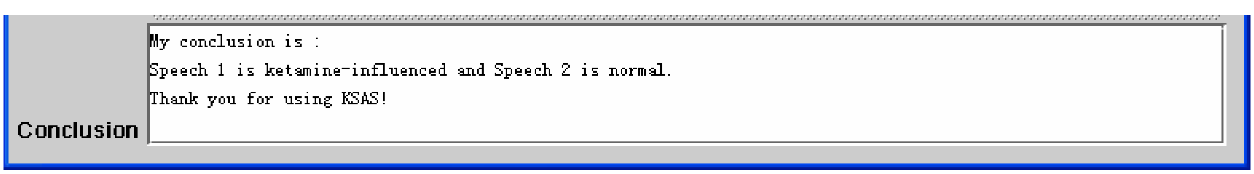
6. Click the "Open .F0 File 2" button and open "sample002.f0." Click the "Open .FRM File 2" button and open "sample002.frm." Click the "Open Transcription 2" button and open "sample002.txt." Now, we should have a window like this.



7. Click the buttons “F0 Basic Stats,” “FRM Basic Stats,” and “Text Stats,” and get the output statistics.



8. Click the “Conclusion” button to get the identification of the ketamine-influenced speech in the conclusion pane similar to the following:



9. Click the “Save Output” button to save an “output.txt” to the folder “Volunteer-A” for future reference. This file will include all the information currently on the window, such as filenames, measures and the conclusion. And that concludes our example.

APPENDIX B

THE KSAS SCORING AND IDENTIFICATION ALGORITHM

```
/*****************************************
* This algorithm returns an int: 1, 2, or 3
* 1. Speech 1 is ketamine-influenced
* 2. Speech 2 is ketamine-influenced
* 3. Cannot decide
*****************************************/
int evaluate(){
//initiation
double score1=0.0, score2=0.0;

//use the five significant measures first
//check percentOfVerbTokens
if(speech1.percentOfVerbTokens < speech2.percentOfVerbTokens){
	score1+=1.5;
}
else if(speech1.percentOfVerbTokens > speech2.percentOfVerbTokens){
	score2+=1.5;
}

//check percentOfNounTypes
if(speech1.percentOfNounTypes > speech2.percentOfNounTypes){
	score1+=1;
}
else if(speech1.percentOfNounTypes < speech2.percentOfNounTypes){
	score2+=1;
}

//check frequency of fillers
if(speech1.filler > speech2.filler){
	score1+=1;
}
```

```
else if(speech1.filler < speech2.filler){
        score2+=1;
}

//check F1 iqrpercent
if(speech1.f1iqrpercent > speech2.f1iqrpercent){
        score1+=1;
}
else if(speech1.f1iqrpercent < speech2.f1iqrpercent){
        score2+=1;
}

//check F1 standard Deviation
if(speech1.f1stdev > speech2.f1stdev){
        score1+=0.5;
}
else if (speech1.f1stdev < speech2.f1stdev){
        score2+=0.5;
}

if(score1>score2)
        return 1;     //speech 1 is ketamine-influenced
else if(score1<score2)
        return 2;     //speech 2 is ketamine-influenced

//if there is a tie, use the less significant measures
else{
        if(speech1.f0stdev > speech2.f0stdev)
                score1+=0.2;
        else if(speech1.f0stdev < speech2.f0stdev)
                score2+=0.2;

        if(speech1.numOfQuestions < speech2.numOfQuestions)
                score1+=0.2;
        else if(speech1.numOfQuestions > speech2.numOfQuestions)
                score2+=0.2;

        if(speech1.freqOfI > speech2.freqOfI)
                score1+=0.2;
        else if(speech1.freqOfI < speech2.freqOfI)
                score2+=0.2;
```

```
        if(speech1.freqOfPos > speech2.freqOfPos)
                score1+=0.2;
        else if(speech1.freqOfPos < speech2.freqOfPos)
                score2+=0.2;

        if(score1>score2)
               return 1;
        else if(score1<score2)
               return 2;

//if there is a tie again, use F3 range as the last measure
        else {
              if(speech1.f3range < speech2.f3range)
                   score1+=0.2;
              else if(speech1.f3range > speech2.f3range)
                   score2+=0.2;

              if(score1>score2)
                   return 1;
              else if(score1<score2)
                   return 2;

//make no decision if there is a tie again
                   else return 3;
        }//else
   }//else
}//evaluate()
```