



Analysis of Voice Signals Change by Voice Modulation Program

*Bong-Hyun Kim

Department of Smart IT, U1 University, 310 Daehak-ro, Yeongdong-eup, Yeongdong-gun, Chungcheongbuk-do, 29131, Republic of Korea

*Corresponding author Email: bhkim@u1.ac.kr

Abstract

Background/Objectives: Voice modulation is used in various fields. Especially, it is widely used in entertainment program for voice tampering to give viewers fun, and voice tampering to guarantee the victim's identity in news. However, in recent years, voice tampering has been exploited for crime. As the information and communication technology in the information society has developed rapidly, the crime using voice modulation is increasing.

Methods/Statistical analysis: Therefore, in this paper, the change of voice signal is analyzed by analyzing both normal voice and modulated voice. For this purpose, general voice was collected using the same place, time, microphone, etc., and a modulated voice was collected by applying a voice modulation program. In addition, various voice analysis parameters such as spectrum, formant, intensity, pitch, pulse, jitter, shimmer, DoVB, and NHR were applied to the study.

Findings: Experimental results show that the difference between the normal voice and the modulated voice is caused by various voice signal analysis parameters due to voice modulation. Especially, in the modulated voice, the spectrum, pitch, and DoVB values were decreased as compared with the general voice. In addition, jitter, shimmer, and NHR values resulted in a result that the modulated voice was higher than the normal voice. There was no significant difference in strength, formant and pulse measurements. Based on the results of this study, it is possible to reflect the changed voice analysis parameters by the voice modulation program.

Improvements/Applications: Voice modulation is useful in various aspects such as fun and identification. However, it has been recently exploited in the same way as voice phishing. Therefore, in this paper, we measured the voice analysis parameters that are changed by the voice modulation program. Based on this, we compared and analyzed the general voice and the modulated voice and extracted the pattern of the voice signal changed by the voice modulation.

Keywords: Voice analysis, Voice modulation, Praat, Voice phishing, Voice parameter.

1. Introduction

The voices have various patterns and are used in many fields by using them. Especially, in entertainment programs voice modulation is used as a tool to give fun and laughter. And in the news and current programs, voice modulation is used as an identification tool of the emergent person. It is also widely used for diagnosis and diagnosis, such as pronunciation analysis and language disorder diagnosis. As such, voice signals are usefully used in various fields.

However, crime reflects society. In other words, as the modern society enters into the 21st century, IT and information and communication technologies are rapidly growing due to the development of electronics industry and semiconductor industry. Finally, as we have been able to share information across borders, crimes such as hacking, malicious code and virus spreading have come to appear, which is a task we must solve to this day. In today's world of smart society through information society, the crime that has been attracting attention due to the development of industries and technologies using smart phones and the increasing use of communication technology is increasing day by day, is "Voice Phishing".

Voice phishing is a combination of 'voice', which means voice, and 'phishing,' which means fraudulent methods of illegally finding and using personal information from financial institutions.

To trick you into deceiving your personal financial information such as passwords, withdraw money, or receive a remittance in the name of a refund, etc. These include fraudulent acts such as impersonating public institutions, financial companies, police, etc., or impersonating relatives' accidents or kidnapping. A new method of fraudulent phone fraud through illegal use of personal information (resident registration number, credit card number, bank account number, etc.) It is a synthesis of voice, private data and fishing. While conventional phishing is a passive way to enter sensitive information via email, voice phishing uses an aggressive method of calling the person who is being committed and requesting remittance or collecting personal information[1-3].

Clear evidence is present in the society of judging crime as DNA and fingerprints, and as a remainder after the events and events of recent microevents have occurred. However, since the voice has been alienated from the evidence of the crime, the crime is being carried out until it exposes its voice more boldly. Especially, if the personal information leaked through phishing has evolved to create the worry of the second crime, we should solve this problem as soon as possible[4-7].

Therefore, in this paper, the change of voice signal is analyzed by analyzing both normal voice and modulated voice. Based on this, we compared and analyzed the general voice and the modulated voice and extracted the pattern of the voice signal changed by the voice modulation. For this purpose, general voice was collected under the same circumstances and conditions, and the collected general voice was converted into a modulated voice by a voice

modulation program. We compared and analyzed the general voice and the modulated voice for the same experimenter, respectively. We analyzed each parameter of voice signal using praat, a voice analysis program, and obtained the results of the study.

2. Research Background and Related Review

2.1. Research Background

Most people can hear other people's voices and judge whether they are people they know. This is because the voice of a person contains information about the personality of the speaker. The personality of the voice is determined by the size, shape, and physical characteristics of the speaker's voice organs, such as larynx, vocal cords, nasal cavities, teeth, etc. In addition, pronunciation, vocabulary, and intonation are influenced by the sociolinguistic environment in which the language is acquired and the individual's utterance habits. As a result, each individual has a unique voice and can be a basis for personal identification by voice[8][9].

Voices can be modulated using a variety of techniques, such as those found in news and current programs, and the modulation technique of these voices uses the nature of the waves. Voice alteration is widely used to protect the identity of the public, such as news, current affairs programs, and crime documentaries. In recent years, it has become popular for entertainment purposes such as entertainment broadcasts and smartphone apps.

While there are voice alterations that are used with this good intent, they are also used for voice phishing crime and can cause a lot of damage. If each person has a unique voice, it is judged that the substance will not lose its quality even if the voice is altered, and still show the owner of the voice. Therefore, by analyzing the difference between the modulated voice and the conventional voice, it is thought that it will help many crime prevention by studying the changing parameter and the unchanging parameter[10-12].

2.2. Related Review

Existing voice modulation studies are related to voice modulation schemes and voice feature changes due to voice modulation. A variety of studies have been conducted in terms of linguistic aspects as a study on the voice modulation method. Voice science research on voice modulation measures good impression in order to persuade people around us and to speak appealing language in everyday life, the voice characteristic of the individual is controlled and used appropriately according to the situation and the environment at that time using the medium of voice. Among the factors involved in communication, things that are done through voice include pronunciation, speed of voice, intensity, intonation, accent, and tone. Such a voice has a greater influence than the meaning of a word because it plays a crucial role in conveying human subtle emotions[13].

From the results of the study, it was found that the favorability of the sound intensity, intensity, and speed of voice using the voice modulating script was higher than that of the original voice in both sexes. The preference for the intensity of the sound is slightly higher than that of the original intensity in men and the intensity of sound is slightly higher in the case of women. And the favorability according to the speed modulation of the horse was higher at the speed of the horse which was slower than that of the male, and the preference was higher at the speed of the horse which was slightly faster for the woman.

A study on the change of voice characteristics by voice modulation has analyzed the effect of helium gas on the voice. In helium gas, it is widely used for voice modulation using the phenomenon that the sound of the voice increases for about 10 to 20 seconds because of the characteristic that the transmission speed is about 3 times faster than that in air. Because of the

characteristics of this helium gas, many people in the TV program as well as the general public are used to induce laughter and interest. In other words, sound is the medium through which air is delivered. At this time, the pitch of sound is determined according to the speed of transmission. When helium gas is consumed, the medium is transmitted through the helium gas from the neck, and the sound of the sound is increased.

The results show that helium gas results in a decrease in pitch value, an increase in jitter value, and a decrease in spectrogram density. It is analyzed that helium gas with low density reduces vibration of vocal cords and disperses voice energy, resulting in a change of original voice pattern, resulting in a lighter voice and a lower pronunciation accuracy.

3. Results and Discussion

In this paper, we have analyzed the parameters that change the original voice according to the voice modulation. Therefore, the original general voice was modulated by voice modulation program and voice analysis was performed. At this time, comparing and analyzing the general voice and the modulated voice, the changing parameter and the unchanging parameter were derived respectively. The experimental procedure is as follows.

- ① Prepare a high-performance sound recorder and a microphone (a built-in microphone and a built-in microphone for smartphone) and a voice modulator (voice modifier .app) as experimental equipment.
- ② In order to measure the change factor by the voice modulation program, the sentence of the greeting form was selected as the free sentence which does not consider specific consonant and vowel.
- ③ In order to reduce the analysis error caused by noise, turn off the power of the peripheral devices and record the sentence in order of 10 persons in total, 5 persons in each male and female in the closed room with windows.
- ④ In consideration of the accurate analysis, the general voice in the state in which the voice modulator is not used and the modulated voice in the state in which the voice modulator are used are recorded separately from each other, and then the normal voice recorded is modified through the voice modulator.
- ⑤ Analyze with a focus on the change of the formant that causes the unique tone of the individual in the general voice and the modulated voice using the voice analyzer (Praat).
- ⑥ Focus on the changes in the intensity of the force applied to the voice from the normal voice and the modulated voice.
- ⑦ Focus on the change of the spectrum that shows the sound wave in the general voice and the modulated voice in the visible form.
- ⑧ Contrast the Voice Report which expresses the numerical value of the pitch value, which means the pitch of the sound, and the jitter value, which means the rate of change of the vocal fold vibration, and the Shimmer value, which means the regularity of the voice amplitude change.
- ⑨ Derive the result based on the analysis.

Experiments were conducted through these experimental procedures. And, in this paper, we compare and analyze general voice and modulated voice by applying various voice signal analysis parameters. First, we performed Formant analysis, which is a voice analysis parameter expressing individual's unique tone in general voice and modulated voice of subjects. A formant is a graph that can clearly show how much energy is present at a certain frequency by analyzing the sound of a person as a frequency. In other words, it can be seen that a specific spectrum appears depending on the difference in vocalization. Thus, the spectral distribution of human sounds is called formants[14-15]. The formant frequency is called the first formant and the second formant from the top of the lower frequency of the figure shown. Figure 1 and Figure 2 shows the waveform and formant frequencies for general voice and modulated voice.

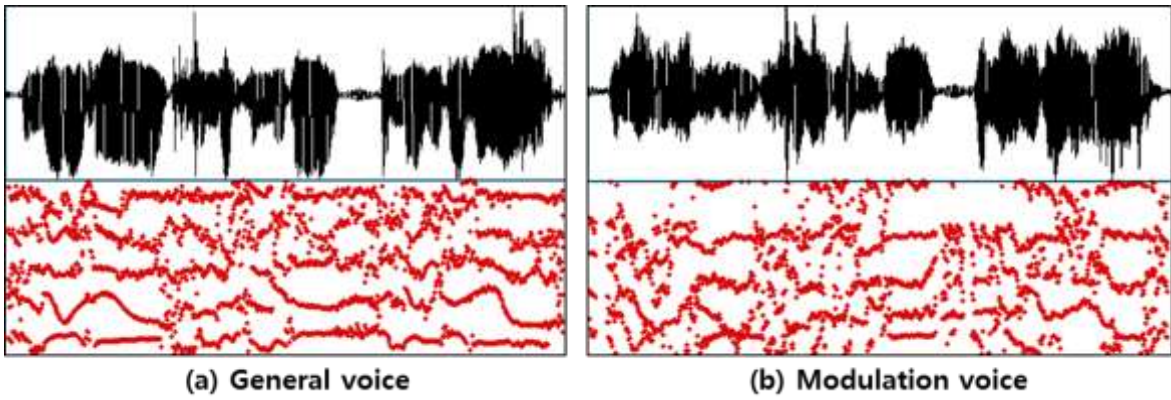


Figure 1: Waveform and formant frequency of male general voice and modulation voice

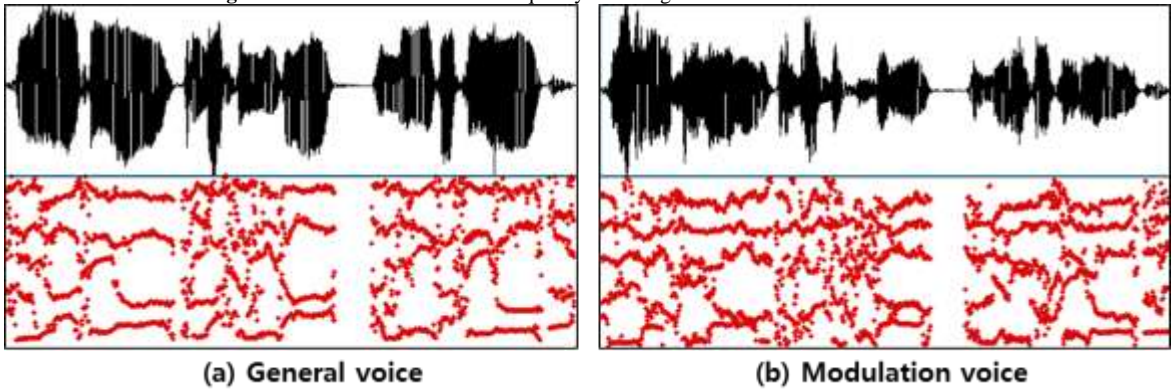


Figure 2: Waveform and formant frequency of female general voice and modulation voice

As a result of the formant frequency analysis, the general voice waveform formed a line style as a whole from one formant to five formant. However, the modulated voice waveform was changed from 1 formant to 5 formant into a dot style instead of a line style, and the line style of a certain section also formed a waveform with a tremble form. These results indicate that the tone changes due to mechanical factors during the modulation process. Next, a intensity analysis showing the magnitude of the voice

energy was performed. Intensity is an parameteral technology for measuring the strength of voice energy in a voice waveform. It means the degree of strength of voice, which is obtained by applying a linear interval to a time point. Intensity refers to the locus of voice energy for linearly distributed time points. Figure 3 and Figure 4 shows the waveform and intensity waveform for general voice and modulated voice.

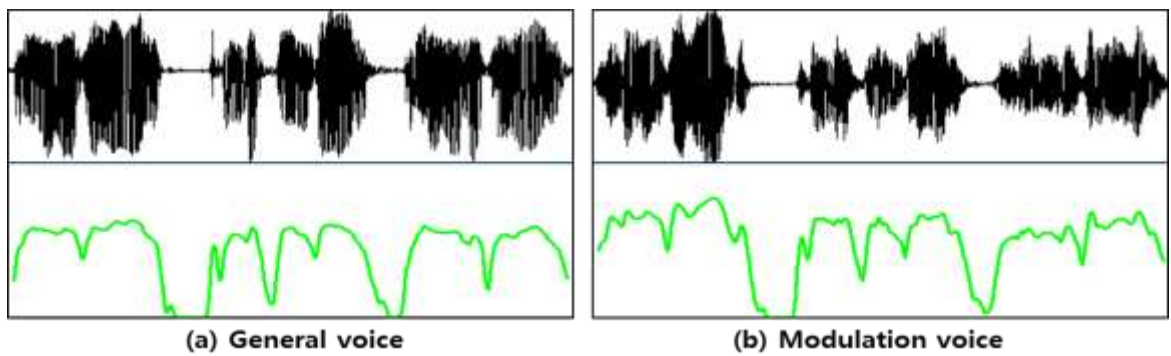


Figure 3: Waveform and intensity of male general voice and modulation voice

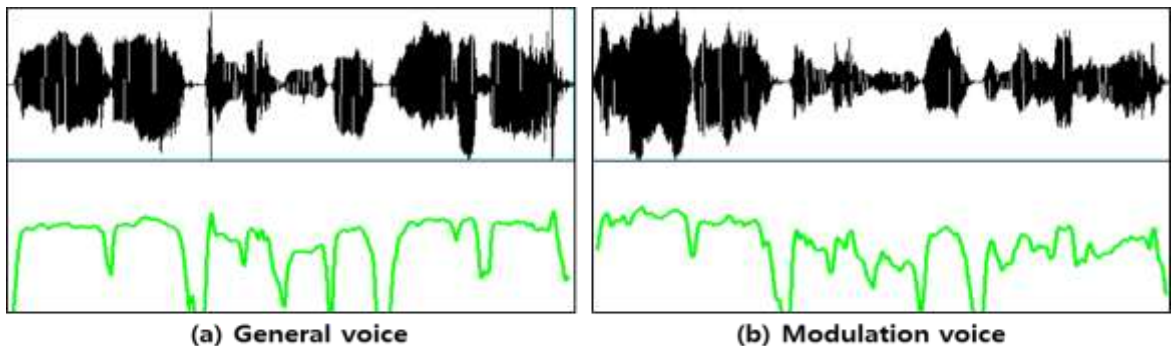


Figure 4: Waveform and intensity of female general voice and modulation voice

As a result of the experiment, it can be seen that the difference between the modulated voice and the normal voice was not large even in the intonation power of the sound, that is, the intonation or the tone, because the intensity means the strength of the voice.

Next, a pitch analysis was performed to measure the vocal fold vibration. Pitch analysis was performed by voice report analysis. Pitch is an analytical parameter that means vibrations of the vocal cords. The height of the voice is expressed acoustically as f_0 . It is usually expressed as the pitch is high and low, keeping in mind the height of the sound that is normally audible. f_0 is also referred to as the fundamental frequency, which depends on the emotional and emotional changes of the speaker. The pitch value indicates how many times the vibrations of the vocal cords are per second, and is affected by the size, length, and mass of the vocal cords[14-16].Figure 5 shows the results of the pitch analysis.As a result of the analysis, it was found that the measured value of vocal cord vibration was decreased in the modulated voice than in the normal voice.

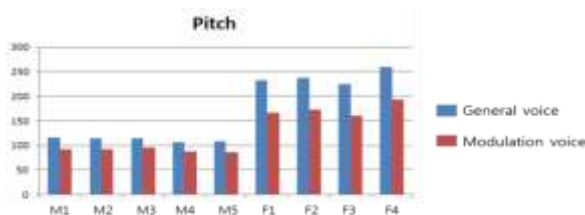


Figure 5: Pitch analysis result

Next, jitter analysis was performed to measure pronunciation accuracy. The jitter analysis measures the instability of the signal, the instability of the signal's amplitude and phase, or either of the short times. In particular, it refers to the left and right vibration state of the signal waveform appearing in the oscillography[15].Figure 6 shows the results of the jitter analysis.As a result, it was found that the pronunciation accuracy was increased in the modulated voice than in the normal voice. That is, the pronunciation of the modulated voice was measured to be inaccurate.



Figure 6: Jitter analysis result

Next, experiments on shimmer analysis were performed. Shimmer shows how regularly the amplitude of each point changes in the voice waveform. Vocalists can make adjustments to gradually increase or decrease the size of the sound by training. In the case of patients such as laryngeal cancer, amplitude of vocal cords is irregular and the amplitude of each vocal point is not constant. Analyze the root-mean-square by squaring the amplitude of each time point.Figure 7 shows the results of the shimmer analysis.As a result of the analysis, the measured value of shimmer analysis increased in the modulated voice than in the normal voice.

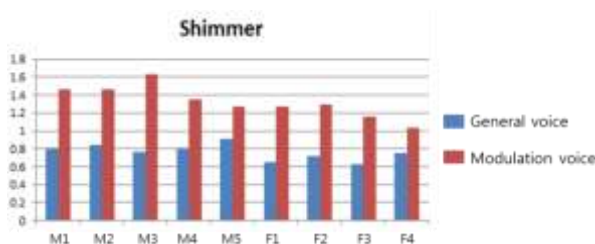


Figure 7: Shimmer analysis result

Next, an analysis for DoVB (Degree of Voice Break) was performed. Degree of voice break is a measure of voice signal breaks. That is, the degree of silence in the voice waveform is measured as a ratio in the entire voice signal.Figure 8 shows the results of the DoVB analysis.As a result of the analysis, the measured value of DoVB analysis decreased in the modulated voice than in the normal voice. This result is considered to be due to the deletion of the silence processing interval.

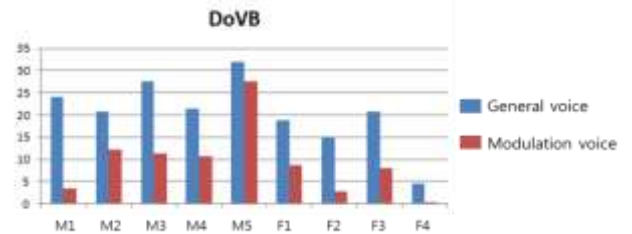


Figure 8: DoVB(Degree of Voice Break) analysis result

Next, an analysis of NHR (Noise to Harmonics Ratio) was performed. The NHR measures the harmonic ratio for noise. That is, the lower the value, the better the measurement of the harmony of pronunciation. Figure 9 shows the results of the NHR analysis.

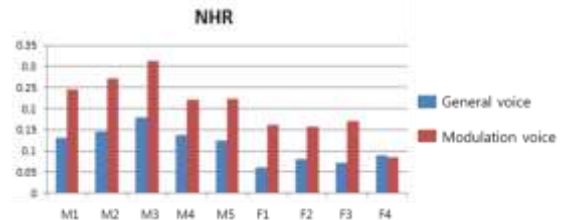


Figure 9: NHR (Noise to Harmonics Ratio) analysis result

As a result, there was no difference in intensity and pulse signal analysis only in comparison and analysis between general voice and modulated voice using various voice analysis parameters. In the remaining voice analysis, the difference between the two groups was confirmed. Especially, pitch analysis and DoVB analysis showed large voice analysis value. However, the jitter analysis, the shimmer analysis, and the NHR analysis showed large values of the modulation voice analysis.

4. Conclusion

Voice modulation is used in various fields. Especially, it is widely used in entertainment program for voice tampering to give viewers fun, and voice tampering to guarantee the victim 's identity in news. However, voice tampering has been exploited for crime. Therefore, in this paper, the change of voice signal is analyzed by analyzing both normal voice and modulated voice. For this purpose, general voice was collected using the same place, time, microphone, etc., and a modulated voice was collected by applying a voice modulation program. In other words, we tried to measure parameters that did not change even when using a voice modulation program.

As a result of the experiment, if the general voice is denatured, it is as if the entire voice is changed, and the personality of the voice is not destroyed. Therefore, it can be found that the owner of the voice which is denatured can be found by contrast with the voice before the denaturalization. This means that not only voice phishing but also violent crime using all means of communication such as kidnapping, explosion warning, intimidation, etc. can be responded quickly.In the future, we will continue to strive to develop technically so that the noise of the denatured voice can be clearly extracted through the filter to identify the owner of the original voice even in the denatured voice.

References

- [1] Aronovitch, C. (1976). The voice of personality: Judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, 99(2), 207-220.
- [2] Brandt, D. (1979). On liking social performance with social competence: Some relations between communicative style and attributions of interpersonal effectiveness. *Human Communication Research*, 5(3), 223-237.
- [3] Bruckert, L., Liénard, J. S., Lacroix, A., Kreuzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the 2006 Conference of the Biological Sciences*, 83-89.
- [4] Buller, D., & Aune, R. (1988). The effects of vocalics and nonverbal sensitivity on compliance. *Human Communication Research*, 14, 301-332.
- [5] Collins, S., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, 65(5), 997-998.
- [6] Ford, W., & Wolvin, A. (1993). The differential impact of a basic communication course on perceived communication competencies in class, work, and social context. *Communication Education*, 42, 215-223.
- [7] Jones, B., Feinberg, D., Debruine, L., Little, A., & Vukovic, J. (2008). Integrating cues of social interest and voice pitch in men's preferences for women's voices. *Biology Letters*, 2(4), 192-194.
- [8] Mehrabian, A. (1969). Some referents and measures of nonverbal behavior. *Behavior Research Methods and Instrumentation*, 1, 213-217.
- [9] Mehrabian, A., & Williams, M. (1969). Nonverbal concomitants of perceived and intended persuasiveness. *Journal of Personality and Social Psychology*, 13, 37-58.
- [10] Miyake, K., & Zuckerman, M. (1993). Beyond personality impressions: effects of physical and vocal attractiveness on false consensus social comparison, affiliation, and assumed and perceived similarity. *Journal of Personality*, 61(3), 411-437.
- [11] Smith, C., Finnegan, E., Karnell, M., & Michael, P. (2005). Resonant voice: Spectral and nasendoscopic analysis. *Journal of Voice*, 19(4), 607-622.
- [12] Street, R., Blady, R., & Lee, R. (1984). Evaluative responses to communicators: The effects of voice rate, sex, and interaction context. *Western Journal of Voice Communication*, 48(1), 14-27.
- [13] Collins, S., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, 65(5), 997-998.
- [14] Caroline H., & Bradley E. (1996). Generating and manipulating emotional synthetic speech on a personal computer. *Multimedia Tools and Applications*, 3(2), 105-125.
- [15] R. Carlson, B. Granström, & I. Karlsson. (1991). Experiments with voice modelling in speech synthesis. *Speech Communication*, 10, 481-489.
- [16] D.W. Massaro & E.L. Ferguson. (1993). Cognitive style and perception: The relationship between category width and speech perception, categorization, and discrimination. *American Journal of Psychology*, 106, 25-49.