A forensic study of whisper and recall

Kyle Smith
University of Victoria
htimsk@uvic.ca

Certain criminal acts, such as fraud and verbal assault, may be of a solely verbal nature. In these cases, it is imperative that prosecutors have an accurate portrayal of exactly what was overheard by ear-witnesses. With aims of maximizing recall by ear-witnesses in judicial situations, this study presents findings based on experimental, comparative recall measures of whispered vs. normal (modal) speech, when the content was originally encountered in a disguised manner (i.e. whispered). Non-word tokens were recorded and acoustically matched in Praat, including 30 target-whisper, 30 test-modal and 30 test-whisper (for a total of 90 tokens). In the first phase of testing, participants heard 10 randomized tokens from the target-whisper bank, to be identified in the second phase via either the whispered or modal prompts. Correct identifications were tallied, and results show a benefit to presenting stimuli in a whisper when originally encountered in a whisper. These findings and applications to ear-witness testimony are discussed, with recommendations toward introduction of whispered stimuli in ear-witness line-ups.

Keywords: whisper; recall; ear-witness

1 Introduction

Ear-witness testimony can often be unreliable. Surprisingly, ear-witnesses who are more certain of their accusations of guilt are more often incorrect (Orchard & Yarmey, 1995). One issue in identification lies with the ease of voice disguise, where simply whispering significantly lowers successful speaker identification: "The easiest and perhaps most common way to disguise a voice is to whisper" (Reich & Duke, 1979, p. 1023). A growing bank of research exists regarding the shortcomings of speaker recognition and identification across voicing styles. Yet, current forensic literature lacks a discussion of witnesses' ability to recall and identify specific content heard in this disguised manner.

A leading researcher in the forensic linguistics field suggests that "for ear-witnesses to be really useful we must find ways of improving their performance for voice identification, and content recall" (Öhman, 2013, p. 9). Despite ear-witnesses' poor performance at identifying disguised voices, common judicial practice still presents only normally-voiced samples to ear-witnesses as a memory aid (akin to eye-witness procedures where a suspect is presented alongside several foils). This study will therefore test the quality of content recall across voicing styles, as to support inclusion of whispered stimuli in suitable police ear-witness line-ups. Research will focus on the recall of content encountered in whispered speech, presented later in normal vs. whispered voicing, across a short timeframe.

2 Background

2.1 Speaker recognition and ear-witness applications

Lisa Öhman perhaps puts it best, stating, “[M]any civil and criminal cases involve testimony regarding statements or content of specific conversations. Furthermore, there are ‘language crimes’ (e.g. verbal sexual harassment, fraud) where the witness’s memory of a conversation is the only available evidence” (p. 9). In these and all cases, it is important that prosecutors have the very best quality of information to come to a correct verdict.

In 1995 Daniel Yarmey developed a comprehensive review of ear-witness speaker identification. He noted that for some crimes the only evidence that may be available and helpful to the courts is the human listener. In such cases where evidence is in short supply, the power of the witness is hard to deny. While this is no place to dissect the judicial system’s practices, there have been many questions related to the reliability of ear-witness testimony, especially when it is well researched that voices are easily disguised by a whisper alone.

Related research in two notable studies involve speaker recognition testing across whisper-whisper tokens, and whisper-normal tokens, among other variables of voice disguise. These studies included presentations of disguised voices, to be identified later among disguised and natural usages, among other variables. Findings from these studies describe that, “the inclusion of a whisper-disguised speech sample in the stimulus pair significantly interfered with listener performance […]” (Reich & Duke, 1979, p. 1028). Secondly, “voice disguise through whispering […] significantly influenced identification performance” (Orchard & Yarmey, 1995, p. 254). This paper utilizes similar methods, focusing on participant’s ability to identify disguised and undisguised tokens. The purpose being to test *recall of content*, rather than *recognition of voices*.

2.2 Acoustics and neurocognitive consequences of whispered speech

In his review of ear-witness identification, Yarmey notes that, “Whispering conceals the most salient characteristics of a voice such as: pitch, inflection, and intonation” (Yarmey, 1995, p. 794). Whispered speech is produced with a more open glottis than normal voicing and with longer syllable durations and stop-closure intervals (Yarmey, 1995, p. 793). In particular, missing from the signal are “pitch and harmonic relationships, with no differential in power between 200_{HZ} and 2000_{HZ}” (Tartter, 1989, p. 1678). These missing characteristics have been positively correlated to the negative impact whispering has on identification of speakers by Tartter, Yarmey, Öhman et al. Since listeners can actively identify and extract useful phonetic information from a whisper, what cues remain for analysis in the utterance?

Seeking more information, in her 1989 article “What’s in a Whisper?”, Vivien Tartter investigated the acoustic characteristics of whispered speech. She revealed that certain cues to identification do remain intact in the signal, such as “nasal resonance and formant dampening, appropriate formant frequencies and transitions, transition durations, and appropriate burst and frication spectra” (p. 1683). In her article, she discusses auditory techniques used by listeners to dissect the deprived signal, such as cognitively determining pitch via the second formant frequency (p. 1683). She further identifies other whispered phoneme identification cues, including “high frequency cues to voicing, such as frication, or burst duration and intensity where low frequency information is lacking” (p. 1683). Despite these remaining characteristics in the whispered signal, Tartter’s research points to a modified analysis of whispered voices compared to that for modal, using reconstructed, or otherwise different cues to phoneme cognition and identification, especially where low frequency cues are lacking.

In a more recent study entitled “Phonetic Feature Encoding in the Human Superior Temporal Gyrus,” researchers were able to isolate the neurological correlates of the entire English phonetic inventory. Using “high-density direct cortical surface recordings,” showed that the superior temporal gyrus relies on “distinct phonetic features for pre-lexical identification of phonemes” (Mesgarani et al., 2014, p. 1). If distinct phonetic features are indeed the neurological correlates of phonemes, it remains to be seen what consequences perception of the whispered speech signal, lacking in those aspects outlined by Tartter, carry forward to the recall of content.

2.3 Prosody, inter-speaker content recall, and non-word inclusion

In 2007, Lisa Archibald and Susan Gathercole tested short-term memory for non-words in school-aged children, finding suggestions that “distinct coarticulatory and prosodic cues may play important roles in recall of multi-syllabic phonological forms” (p. 604). Aligned with this finding, in 1988 John Mullennix, David Pisoni, and Christopher Martin showed that simply exchanging speakers in the test phase had negative effects on recall of items. Their study shows that the effects of talker variability are “more robust and less dependent on task than word frequency or lexical structure” (p. 375). They go on to propose that “information about the talker’s voice is intimately related to early perceptual processes that extract acoustic-phonetic information from the speech signal” (p. 375). Considering the results of these studies, it seems content recall is affected by prosodic cues and voice characteristics, similar to those in recognition outlined above. Research presented below will consider these findings with necessary methodological adaptations to Orchard and Yarmey’s 1995 research on recognition.

2.4 Research question and hypothesis

Q) Is recall of target items affected by the use of different voice qualities across stimulus pairs (whisper-whisper vs. whisper-normal)?

H_1) Given the acoustic differences between whispered and normal voicing discussed above, it is predicted that cross modal recall will be weaker (whisper-normal) than same mode recall (whisper-whisper).

4 Methodology

4.1 Participants

20 undergraduate students from the University of Victoria were tested. The participants were evenly split into test (whisper-normal), and control (whisper-whisper) groups. All were native English speakers with no reported hearing deficiencies and normal or corrected to normal vision.

4.2 Stimuli

A total of 90 bisyllabic, possible English non-word tokens (e.g. 'artson', 'juber') were recorded in Praat, using a Headrush USB headset. Tokens followed English phonotactic constraints and were developed at random for the purposes of this experiment. The 90 tokens were comprised of 30 "target-whisper", 30 "test-normal" and 30 "test-whisper" stimuli, which were recorded separately from the "target whisper" tokens. Each set was composed of the same 30 non-words. The non-words were all initial-syllable stressed for regularity. All tokens were provided by the same male speaker and acoustically matched in Praat for duration, relative amplitude, and volume to minimize external differences in the testing. Tokens were as clear as possible, containing no background noise or auditory glitches that may have aided in recall. As non-words present "a relatively pure measure of phonological short term memory," lexical and word frequency effects should be eliminated via this method (Archibald & Gathercole, 2007, p. 602). A text list of the 30 non-word stimuli is presented in Appendix A. Audio tokens were loaded onto a Samsung Galaxy S3 phone, and presented at normal listening volume with headphones via the default Samsung player. This maximized portability and greatly aided in recruiting participants. Modal-modal stimuli were not presented based on the small scale of this research.

4.3 Procedure

Participants were randomly assigned to groups (whisper-whisper, whisper-normal), and were individually presented 10 randomized "target-whisper" tokens, a subset of the 30 test words, heard in whispered voicing (See Appendix A). Each participant heard a different, randomized subset of these target items. Randomization was accomplished via the shuffle function on the player. Recall testing presented all 30 tokens, in whispered or normal speech as per their group. Again, whisper-whisper trials made use of different whispered recordings for presentation and test phases (i.e. target-whisper, test-whisper) to equalize the test groups in terms of the number of times they encountered the stimuli.

After randomized target presentation, testing consisted of the participants receiving an ordered test page on which to indicate the tokens they recalled. At this point either whispered or normally voiced prompts were presented as per the participant's group. Audio tokens were presented in the same order as listed on the test page (see Appendix A), as to minimize participant's time spent searching for already randomized stimuli. A one-second delay between tokens was included for participant's processing.

4.4 Analysis

Participant scores were based on accurate identification of targets from the presentation phase, with scores out of 10 for each trial. Correct identifications were each worth one point, where incorrect identifications (akin to a wrongful accusal) deducted a point from the score. Simple descriptive statistics of the between groups, dependant variable scores (*/10) such as average, median, and standard deviation were calculated in Excel. Results are presented below.

5 Results

As predicted, testing showed an advantage to recall when stimuli were of the same modality, "whisper-whisper" (mean = 4.5, SD = 2.8), than when of different modality "whisper-normal" (mean = 3.1, SD = 2.12). As detailed below in Figure 1, average scores were higher, though more variable in the whisper-whisper condition. The overall trend shows the linear averages of participants' scores. Complete raw data are presented in Appendix B in table format.

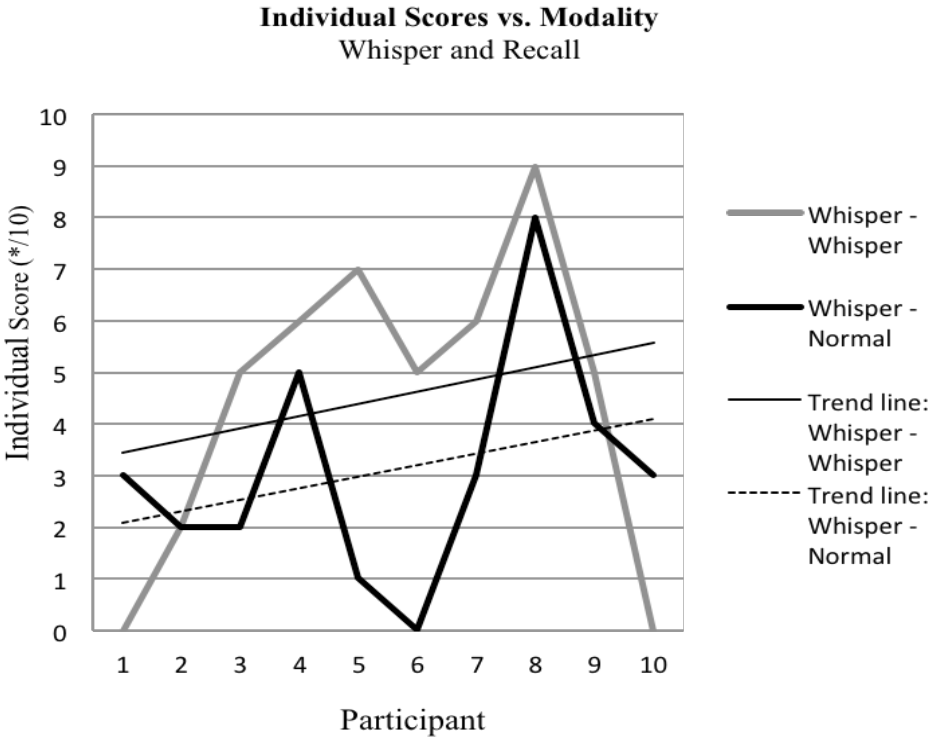

Figure 1: Modality vs. Individual Score of each of 20 the participants (10 per group). Whisper-Whisper is presented in grey, with Whisper-Normal in ***black****. Linear averages appear in solid and checked lines respectively.*

6 Discussion

These results are aligned with findings from voice recognition studies and relatable studies on item recall. As with Orchard and Yarmey, Ohman, and Tartter's findings discussed above voice disguises, such as whispering, have measurable impact on listener's ability to connect utterances made by the same speaker. It makes intuitive sense that the more alike a prompt is, the better the recall will be. As such, this study supports the inclusion of whispered stimuli in applicable forensic scenarios where the accused had whispered in the original encounter. By extension, a reasonable assumption can be made that other common voice disguises (eg. harsh, creaky, etc.) should be mimicked in any recall aids. Future testing is recommended to verify this assumption regarding item recall, though similar studies on recognition have already reached similar conclusions.

The direct cause of the observed disparity in average scores is only speculative at this point, though inherent acoustic differences between whispered and normal speech certainly play a role.

7 Conclusion

Same and cross-modal recall test scores were compared. Findings showed that a benefit exists to providing ear-witnesses with as similar a reconstruction as possible when aiding recall in judicial applications. The observed data from this study, as well as those of the experiments discussed above, support this conclusion. As no financial, ethical or other considerations are affected by this method, investigators would do well to adapt their methodologies to reflect the careful findings of the scientific community.

Acknowledgements

Thanks to Dr. Sonya Bird for her time, patience, and valuable input throughout the process of putting this study together. Further thanks go to the whole of the Ling 486 class for their enthusiasm and aid. Finally, thanks to all the participants who freely devoted their time, without whom there would be no study to present.

References

Archibald, L. M. D. & Gathercole, S. E. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics,* 28(4), 587-606.

Boersma, P. & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.3.71, retrieved 9 February, 2014 from http://www.praat.org/

Mesgarani, N., Cheung, C, Johnson, K. & Chang, E. F. (2014) Phonetic feature encoding in human superior temporal gyrus. *Science.* doi: 10.1126/science.1245994

Mullennix, J. W, Pisoni, D. B. & Martin, S. C. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America,* 85(1), 365-378.

Öhman, L. (2013). *All ears: Adults' and children's ear-witness testimony*. (Doctoral Dissertation) Retrieved From http://hdl.handle.1101-718 Xnet/201101-718X77/32014. Accessed February, 2013

Orchard, T. & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, 9(3), 249-260.

Reich, A. R. & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *The Journal of the Acoustical Society of America,* 66(4), 1023-1028.

Tartter, V. C. (1989). What's in a whisper? *The Journal of the Acoustical Society of America,* 86(5), 1678.

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law,* 1(4), 792-816.

Appendix A

30 token non-words (in test order). Non-word test stimuli were presented in this order and according to modality. Participants checked off the stimuli that they recalled from the presentation phase on the test page. Scores were out of 10, with incorrect responses deducting a point from the score.

1. Flondo
2. Artson
3. Classit
4. Govish
5. Daxon
6. Badan
7. Rontol
8. Wabled
9. Brofson
10. Pendle
11. Aidom
12. Dooted
13. Tiskler
14. Banton
15. Gassive
16. Ostush
17. Waltet
18. Scraffle
19. Tromson
20. Scouble
21. Dufsit
22. Guitan
23. Ipsit
24. Popten
25. Maendle
26. Rissle
27. Embrit
28. Nersten
29. Haltred
30. Juber

Appendix B

T*able 2: Dependant variable scores for each participant (1-10), arranged by test group. Average, median, and standard deviation of the data points are presented at the bottom*

	Recorded Data	
Participant	Whisp-Whisp	Whisp-Norm
1	0	3
2	2	2
3	5	2
4	6	5
5	7	1
6	5	0
7	6	3
8	9	8
9	5	4
10	0	3
Average	4.50	3.10
Median	5	3
Standard Deviation	2.80	2.12