LEARNING TO MATCH FACES AND VOICES

by

Meredith Davidson

A Thesis Submitted to the Faculty of

The Charles E. Schmidt College of Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

Florida Atlantic University

Boca Raton, Florida

August 2010

LEARNING TO MATCH FACES AND VOICES

by

Meredith Davidson

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Elan Barenholtz, Department of Psychology, and has been approved by the members of her supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Master of Arts.

SUPERVISORY COMMITTEE:

Elan Barenholtz, Ph.D.
Thesis Advisor

Alan W. Kersten, Ph.D.

David J. Lewkowicz, Ph.D

David L. Wolgin, Ph.D.
Chair, Department of Psychology

Gary W. Perry, Ph.D.
Dean, The Charles E. Schmidt College of Science

Barry T. Rosson, Ph.D.
Dean, Graduate College

May 14, 2010
Date

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Elan Barenholtz, for accepting me into his lab and providing continuous intellectual guidance throughout the course of my Masters of Arts Degree. I would also like to thank my family and my friends who gave me their unconditional support throughout this process.

ABSTRACT

Author:	Meredith L. Davidson
Title:	Learning to Match Faces and Voices
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Elan Barenholtz
Degree:	Master of Arts
Year:	2010

This study examines whether forming a single identity is crucial to learning to bind faces and voices, or if people are equally able to do so without tying this information to an identity. To test this, individuals learned paired faces and voices that were in one of three different conditions: True voice, Gender Matched, or Gender Mismatched conditions. Performance was measured in a training phase as well as a test phase, and results show that participants were able to learn more quickly and have higher overall performance for learning in the True Voice and Gender Matched conditions. During the test phase, performance was almost at chance in the Gender Mismatched condition which may mean that learning in the training phase was simply memorization of the pairings for this condition. Results support the hypothesis that learning to bind faces and voices is a process that involves forming a supramodal identity from multisensory learning.

DEDICATION

This manuscript is dedicated to my family, particularly to my late father, Bill, who always encouraged me in my academic endeavors as well as provided support during any of my trying times. Through his belief in me, I have been able to achieve my goals.

LEARNING TO MATCH FACES AND VOICES

LIST OF FIGURES

INTRODUCTION

After talking to someone on the phone that one has not met and later meeting that individual in person, people are sometimes shocked to see that the individual looks nothing like they had imagined they would from the sound of their voice. Clearly, people have preconceived notions of what people should look like based on the sound of their voice. Where does this expectation come from? Although voices and faces can be heard or seen alone, they are usually presented together and are both part of a single identity to which they are bound. Thus learning an identity is inherently a form of multimodal association. However, little is known about how such multimodal representations are developed. One hypothesis is that faces and voices are combined into a single 'supramodal' identity which makes them retrievable (Mesulam, 1998). According to this view, faces and voices are not simply independent signals that are associatively paired based on repetitive conjunction, but instead represent different properties of the same individual. In person recognition, parts of an individual such as their face and voice are assembled in the mind into representations of that individual. The supramodal representation of an identity can be seen as a higher level of person identification because sensory cortices are already reactive to the crossmodal information that comes from the face and the voice (von Kreigstein & Giraud, 2006). An alternative hypothesis is that individuals are learning to pair faces and voices together because faces and voices are independently important and present overlapping information, or possibly that faces and voices are bound together in a similar fashion to other arbitrary information.

The Information In Faces and Voices

Although faces and voices are both aspects of identity, they carry different and sometimes overlapping information. For example, both faces and voices convey information about age, gender, size, and ethnic group (Kamachi, Hill, Lander & Vatikiotis-Bateson, 2003; Lachs & Pisoni, 2004b; von Kriegstein & Giraud, 2006). Voices alone are often sufficient to convey attributes of the speaker to the listener, including affective information that is informative about the speaker's internal state such as prosody, which conveys emotion through amplitude, vocal pauses, and variation of the fundamental frequency (Belin, Fecteau, & Bedard, 2004; Krauss, Freyberg, & Morsella, 2002). In addition, voices carry information about visible physical attributes. For example, Krauss et al. (2002) found that not only were participants able to identify speakers based on age and size in general, but they also were able to provide accurate estimates of both the speakers' height and weight. Other more controversial findings include that listeners can identify a speaker's socioeconomic status, racial group, and attractiveness (Hughes, Dispenza, & Gallup, 2004; Kreiman, 1997; Zuckerman, Miyake, & Elkin, 1995). As with recognition of voices, faces alone also contain much information about an individual based on both static and dynamic face stimuli. Both familiar and unfamiliar faces can provide information about the individual's gender, race, and age (Cloutier, Mason, & Macrae, 2005, Eberhardt, 2005; Macrae, Quinn, Mason, & Quadflieg, 2005), emotions (Calder & Young, 2005; Horstmann & Ansorge, 2009), and attractiveness (Zuckerman, Miyaki, & Elkin, 1995). More controversial studies have also found that faces can provide information for intelligence, competence, and dominance as well (DeBruine, 2005; Todorov, Mandisodza, Goren, & Hall, 2005; Zebrowitz & Collins, 1997). Face processing of familiar individuals is also highly efficient and adults can identify a person's name to their face as fast as they can identify that the person is human (Schwarzer & Leder, 2001).

Neural Effects of Faces and Voices

Multimodal integration can also be evaluated by addressing the neural effects that this integration has on the brain. Specifically, for vocal information, regions located along the superior bank of the Superior Temporal Sulcus (STS) have shown a greater response to vocal sounds than to non-speech related sounds which provides evidence that voice perception involves specialized mechanisms in the brain due to areas in the brain showing voice-sensitive cortical activity (Belin et al., 2004; Hickok & Poeppel, 2000; Zatorre & Binder, 2000). Specifically for facial recognition neuroimaging studies, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have found that in the fusiform gyrus is a face-selective region and is called the Fusiform Face Area (FFA) (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Gauthier, Tarr, Aanderson, & Skudlarski, 1999).

Integrating Faces and Voices

While each modality provides information independently and distinct neural mechanisms are devoted to processing them, there is evidence that these modalities are integrated as well. There is a substantial literature demonstrating the integration of face and voice information in the perception of spoken speech. One of the most prominent examples is the so-called 'McGurk effect' (McGurk and MacDonald, 1976). In this phenomenon, the observer sees a video clip of a person pronouncing one phoneme while hearing an audio clip of the person pronouncing a different phoneme, but the observer perceives a third, intermediate phoneme. This phenomenon indicates that visual and auditory stimuli are directly integrated during speech perception (McGurk & MacDonald, 1976).

A number of studies have examined people's ability to infer speaker's faces from their voices and vice versa. Kamachi and colleagues (2003) examined whether participants were able to match a video of an unfamiliar face to an unfamiliar voice. The participants were shown either a face or a voice (dependent on which condition they are in) in the first phase, and then they were presented with two voices or faces in the second phase. The participants were asked to choose which stimuli from the second phase belonged to that presented in the first phase. Participants were able to correctly match the identity of unfamiliar people across visual and auditory modalities, but only when they were shown video stimuli presented in the right direction. When stimuli were temporally reversed or when the visual images presented were only static images, performance was at chance. Interestingly, the matching effect does not seem to be dependent on language, as English participants were able to match Japanese stimuli in further preliminary experiments by Kamachi and colleagues. Instead, the authors conclude that participants were sensitive to the matching temporal information in the visual and auditory stimuli, specifically for normal spoken language.

In a similar study, Lachs and Pisoni (2004) found that information about the source of utterances was available through visual and auditory displays of speech despite removal of the f0 frequency from the audio track and even for single spoken words. This illustrates that even without the f0 frequency, a cue that has traditionally been used to demonstrate vocal identity; the acoustic signal still carries the necessary information for listeners to match the voice to the face. Like Kamachi and colleagues, (2003) the cross-modal information was not available with only static images or if the stimuli were temporally reversed.

Lander and colleagues (2007) researched whether changing sentence content or manner of speaking impairs matching the identity of the face and voice of an unfamiliar person, even when speaking in an unfamiliar language. In both conditions, participants viewed a silently moving face and then chose from two voices. In the sentence-content condition, the moving face was producing a different sentence than the one the participants heard; while in the manner-of-speaking condition, the moving face was producing a sentence in statement form, but the voice was producing a sentence in question form or vice versa. Lander and colleagues found that changing manner impaired the ability for participants to match the faces to voices, but changing sentence content did not.

One possibility in explaining the previous findings is that human speech provides a rich temporal sequence that can be matched based on temporal synchrony alone, and does not provide evidence for specialized integration of faces and voices. However, Tuomainen and colleagues (2005) tested whether people could correctly match a speaker to a corresponding sine wave speech stimulus (SWS). SWS is created by only retaining the course-grain changes in the vocal spectrum and filtering out the fine-grain acoustic cues, which leaves the speech to sound like patterns of bells and whistles if not explained that the listener is hearing a synthesized sentence. Participants, who at first were unaware of the SWS being speech-like, did not integrate the auditory stimuli with the faces they saw. However, when they were made aware of the speech-like nature of SWS, they were able to integrate the same auditory stimuli with the faces. The authors concluded that when the SWS was perceived as non-speech, despite the presence of the same temporal information, the auditory and visual stimuli were processed independently and did not form into a multisensory object. However, when perceived as speech, specialized mechanisms for binding the features into a multisensory object were engaged.

Binding Faces and Voices

The previous research concerns the interaction between visual and auditory information in perceiving spoken speech as well as the ability to infer a face from a voice while observing spoken speech. More recently, researchers have begun looking at how multimodal information from faces and voices is integrated into an identity.

The process of integrating identity through faces and voices seems to occur early in development. Brookes and colleagues (2001) found that three month old infants looked longer at a novel combination of voice/face pairings than voice/face pairings with which the infants had been familiarized.

In an adult study concerning the representation of already learned faces and voices, Schweinberger, Roberston, & Kaufmann (2007) dubbed familiar or unfamiliar voices onto moving faces articulating the same sentences. They then tested the ability of observers to identify the voice as familiar or unfamiliar depending on whether they were paired with the correct familiar face, an incorrect familiar face or an incorrect unfamiliar face. Overall, they found that when the familiar voices were shown together with dynamically presented (i.e. dubbed), unfamiliar faces there was a cost in identifying the voice as familiar. However, no such deficit was found when the faces were shown statically or when paired with an incorrect familiar voice. These findings suggest that even though the task only required recognizing the voice alone, the presentation of the unfamiliar dynamic face stimuli was automatically integrated and interfered with the task.

While this study does examine the ability of people to integrate audiovisual information into an identity, it deals with face and voice binding that has already taken place (when the familiar identities were learned) and not the process of learning to bind faces and voices. One study that has looked at active face-voice binding is von Kreigstein and Giraud (2006), who used both neurological and behavioral methods to examine face and voice binding and also compared face and voice binding to other forms of multimodal association such as binding ring tones to cell phones.

Using neurological methods, Von Kriegstein and Giraud (2006) have examined how the perception of faces and voices in conjunction are activated in the brain. They found that multisensory association, the ability to combine features such as faces and voices in the brain, is generated by brief learning of voice and face pairings. Additionally, this study examined the neural effects as well as the behavioral effects of the multimodal combinations. The neural impact according to the von Kriegstein and Giraud study is evident in the increased activity within the anterior temporal cortex when participants matched voice-face and voice-name as opposed to ring tone-cell phone, and the activity in the anterior temporal cortex was higher the more rapidly participants responded.

After voice-face learning, a direct connection developed between the fusiform face areas (FFA) and the temporal voice areas (TVA). Von Kriegstein and Giraud found that the anterior temporal cortex is an important area of the brain for person recognition at the advanced processing stage, the supramodal level of identification. Previous studies had suggested that the anterior temporal region is heavily involved in multimodal representations of individuals, and that stimuli such as faces and voices involve other areas of the brain prior to accessing a common neural pathway which converges to personal identity (Gorno-Tempini et al., 1998; von Kriegstein & Giraud, 2006). Overall, stronger results were discovered for voice-face recognition than for voice-name recognition and pairing voices and faces together creates robust multisensory associations.

In a set of complementary behavioral studies von Kriegstein and Giraud trained one group of participants to associate voices with faces, while the other group learned to pair voices with names. They found that associating faces with voices improved later recognition when the voice was presented alone by about 14%, while associating a name with a voice only improved later speaker recognition by 5%. Longer response times and higher error rates were also noted for the voice-name condition as compared with the voice-face condition. The researchers suggest that the multimodal stimuli form a particularly robust representation. In order to determine whether faces and voices are special, as opposed to simply being multimodal, von Kriegstein and Giraud also had both the group that learned to match faces with voices and the group that learned to match names with voices learn to associate certain ring tones with pictures of cell phones. The condition of ring tones to cell phones was used because they are similar to faces and voices in that they are two modalities coming from the same source. The results showed an association of ring tone to cell phone improved later ring tone recognition by only 5%, similar to the voice-name pairing condition, and much less compared with 14% improvement for the face-voice paired condition. These results indicate that the exposure to the voice-face pairing forms a multimodal representation that is more robust than those formed by arbitrary association.

Overall, the above studies demonstrate that people are sensitive to the temporal synchrony in faces and voices and that their integration can lead to a more robust representation of identity. However several essential questions remain concerning the process and circumstances under which this binding occurs. One important question is whether binding faces and voices occurs because both faces and voices are independently important and carry overlapping information, unlike, for example, cell phones and ring tones. Alternatively, perhaps face and voice integration represents a specialized process specifically devoted to forming a ‘supramodal’ identity of an individual and is distinct from other types of associations.

Present Research

To test this, the current study examines whether forming a single identity is crucial to learning to bind faces and voices, or if people are equally able to bind faces and voices without tying this information to an identity. If being able to form an identity is crucial to binding faces and voices, then the prediction is that people would be less able to bind faces and voices in circumstances where the pairing of faces and voices is more difficult. In particular, we compare learning under conditions where the faces and voices are ‘congruent’— both of the same gender —vs. learning when they are ‘incongruent’—of the opposite gender. The experimental hypothesis of this research is that face/voice pairings that are incongruent are more difficult to learn because they are not bound into a single identity as compared with congruent pairs that may be bound into an identity.

To examine this, participants in this proposed research study performed a task in which they learned to match pictures of faces with recordings of people speaking a sentence. They then performed a test phase to determine how well they learned to pair the faces and voices using new sentences they did not encounter during the learning phase. There are three different conditions, 1). True voice: in this condition the faces and voices have been recorded from the same individual, 2). Gender-matched: in this condition, faces and voices are not the true faces and voices, but have been matched according to gender and 3). Gender-mismatched: in this condition, the faces and voices are of opposite gender.

The reasoning behind these conditions is as follows: if forming an identity is a specialized process particular to binding faces and voices, then performance in these tasks should be better when the faces and voices are of the same gender (Conditions 1 and 2) and may therefore be combined into a single identity as compared with the gender-mismatched condition (Condition 3), where identity formation will be impaired if not impossible. If, on the other hand, binding faces and voices is simply a form of associative pairing, and does not involve forming an 'identity', then they should be able to perform equally in all conditions.

Within the same-gender conditions (1 and 2) we expected, based on previous results, that people cannot infer voices from static pictures of faces. However, it is possible that differences will emerge in this binding task that was not present in earlier studies. In any case, it is necessary to include the same-gender condition to exclude the possibility that subjects are inferring faces from voices based on the ability to guess rather than based on training.

The task participants performed contained both a Learning phase and a Test phase. The learning phase contained stimuli consisting of different speakers all repeating the same sentence with the task of learning to match each picture with a particular voice recording of that sentence. In the learning phase, participants received feedback as to whether they are correctly matching the face to the voice. In the Test phase participants were asked to match the same faces and voices they have encountered in the Learning phase for recordings of new sentences they had not heard during the Learning phase and without feedback. Performance is measured both in the learning and the test phase. The learning phase measures the ability to learn a specific auditory-visual stimulus pair, whereas the test phase is examined to determine the ability to generalize to new stimuli based on learning to match the faces with voices, not just particular auditory stimuli. Thus, an important difference between the two phases is that it is possible that learning in the first phase is based just on associative pairing of a face and a voice (in other words learning various cues that the faces and voices provide as opposed to truly learning to bind the faces and voices), while the test phase measures the participants' ability to truly learn the match of a face and a voice.

METHOD

Participants

Participants are Florida Atlantic University undergraduate students participating for course credit. The experiment uses a Between-Subject design with twenty-five different students participating in each of the three conditions.

Stimuli

Stimuli consist of photographs and voice recordings of 8 Caucasian females and 8 Caucasian males ranging in age from 18-30. Each individual has been photographed and recorded speaking three sentences: 1) “There are clouds in the sky”, 2) “The boy took his sister to the park”, and counting from one to five. All clips display the head and shoulders of the person from a frontal viewpoint.

Before the beginning of the experiment, each of the 16 face images was matched with a single recorded voice as the ‘pair’ to be learned by the subject. There are three types of pairs, comprising the three experimental conditions: True Voice, Gender Matched and Gender Mismatched. In the True Voice condition, the pairs consist of the face and voice of a single individual (that is, the face and voice truly match) while in the two additional conditions the pairs consist of an individual’s face paired with a recording of one of the other recorded voices: In the Gender Matched condition each picture is uniquely paired with one randomly chosen voice of the same gender, with the constraint that it not be the true matching voice. In the Gender Mismatched condition, each of the female faces is paired with a single randomly chosen male voice and vice versa.

Procedure

The procedure is nearly identical across all three conditions. Participants have first been instructed that that they will be performing a task in which they must learn to match faces and voices. In the gender-mismatch condition, subjects were explicitly informed that the faces and voices do not belong to the same person.

On each trial, the participants were first presented with a voice recording of one of the three recorded sentences, and at the same time four faces were presented on screen in a numbered row in random order. One of the four faces is be the ‘match’ to the voice, as determined prior to the experiment as described above. The subjects were instructed to choose which of the four faces is matched with the voice. An incorrect response results in a low beeping sound, with the correct selection flashing once before the screen is refreshed. The learning phase consists of six blocks using one of the three recorded sentences. Each experimental block consists of four sequences, each consisting of a group of four female or male voices played in sequence across four trials. Within a given sequence, the same four faces appear as the target choices from which subjects must select as a match to the voice (i.e. each face appears on every trial of that four-trial sequence and *only* in that sequence).

After the participant had completed the learning phase, the test phase began. The participant was presented with a voice recording of a different sentence (but from the same speakers as in the Learning phase) that was not used during the learning phase. The participant was then asked to match the voices to the faces that they previously learned in the learning phase without receiving feedback (in order to ensure that no learning takes place during the test phase). The participants were then tested on each face and voice that they had learned during the learning phase (16 faces) on the two new sentences, for a total of 32 test trials for each subject.

RESULTS

Performance was analyzed in terms of percentage correct across the True Voice condition, the Gender Matched condition, and the Gender Mismatched condition. Performance was analyzed in both the Training phase as well as the Test phase. The prediction is that congruent pairings (i.e. same gender) would be easier to learn than incongruent ones (i.e. opposite gender). This hypothesized benefit could appear in the Learning Phase—in the form of quicker learning or higher ultimate performance—or in Test Phase or in both phases.

Training Phase

Figure 1 shows the difference in performance between all three conditions in the training phase. Performance in the True Voice condition and the Gender Matched condition are nearly identical. This confirms earlier findings that people do not discriminate between unfamiliar voice and face pairings when the stimuli are static (Kamachi et al., 2003). Figure 1 also shows that while learning did occur in the Gender Mismatched condition, performance in this condition was much poorer than performance in either of the other two conditions.

Specifically, in the training phase, a one-way analysis of variance (ANOVA) examined the effects of learning between the three groups of face and voice pairings. Learning differed significantly across the three groups, *F*(2,74)=28.14, *p*<.01. LSD Post-hoc comparisons of the three groups indicate that learning in the true voice condition (Condition 1) and learning in the gender matched condition (Condition 2) differed significantly from learning in the gender mismatched condition (Condition 3), with p<.01. No significant difference was found between learning in the true voice condition and the gender matched condition. The analysis of the training phase indicated that quicker learning occurred with the True Voice and the Gender Matched conditions than with the Gender Mismatched condition (see Figure 1).

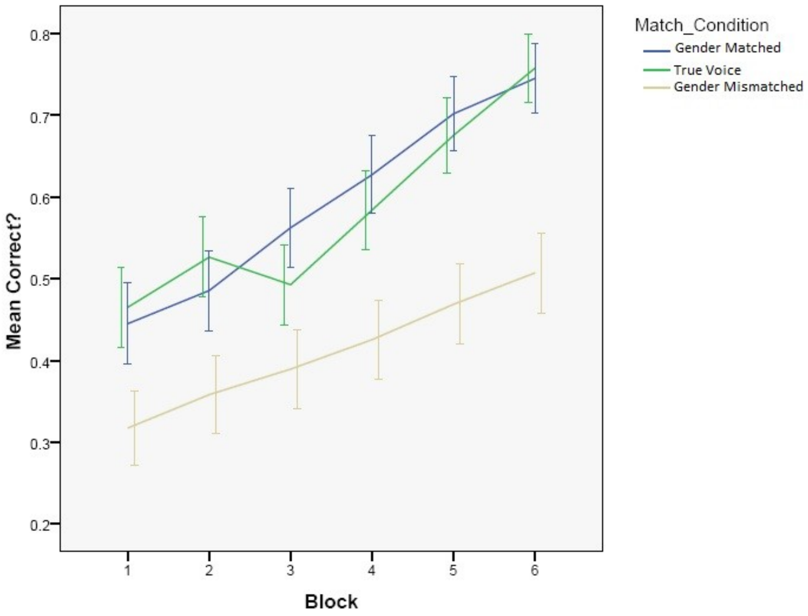


Figure 1. Difference between Conditions 1-3 in Training Phase

Test Phase

Figure 2 shows the difference in performance in the test phase between all three conditions. Performance in the True Voice condition and in the Gender Matched condition are shown to be identical, while performance in the Gender Mismatched condition is much poorer and nearly at chance.

In the test phase, a one-way ANOVA examined the effects of performance between the three groups of face and voice pairings. Performance differed significantly across the three groups, *F*(2,74)=9.58, *p*<.01. LSD Post-hoc comparisons of the three groups indicate that performance in the True Voice condition (Condition 1) and the Gender Matched condition (Condition 2) differed significantly from performance in the Gender Mismatched condition (Condition 3), with p<.01. No significant difference was found between performance in the true voice condition and the gender matched condition. Results indicated higher ultimate performance in the True Voice and Gender Matched conditions.

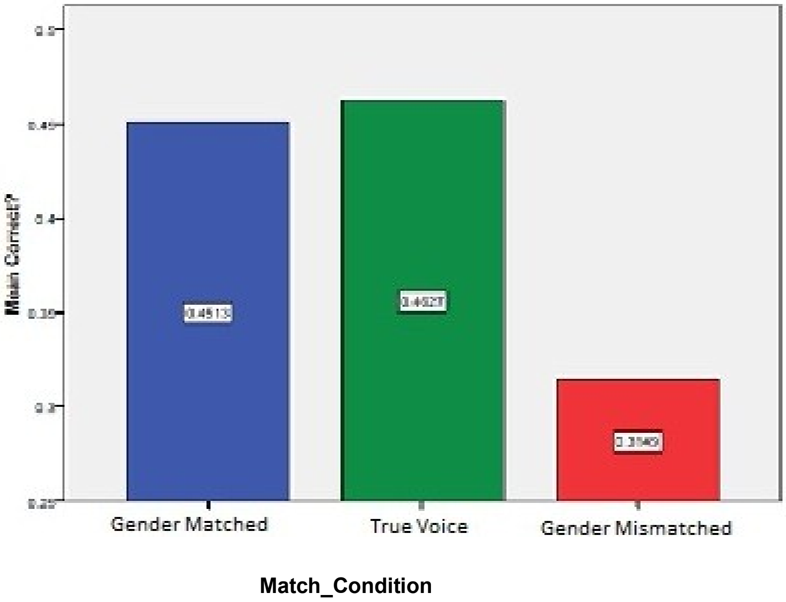


Figure 2. Difference between Conditions 1-3 in the test phase.

Overall Performance

Overall, differences in performance were found in the both the training and test phases between the True Voice and the Gender Mismatched conditions, as well as between the Gender Matched and Gender Mismatched conditions. Performance was lower in the Gender Mismatched condition in both the learning and the test phase; specifically both learning and ultimate overall performance was lower in this condition (see Figure 3).

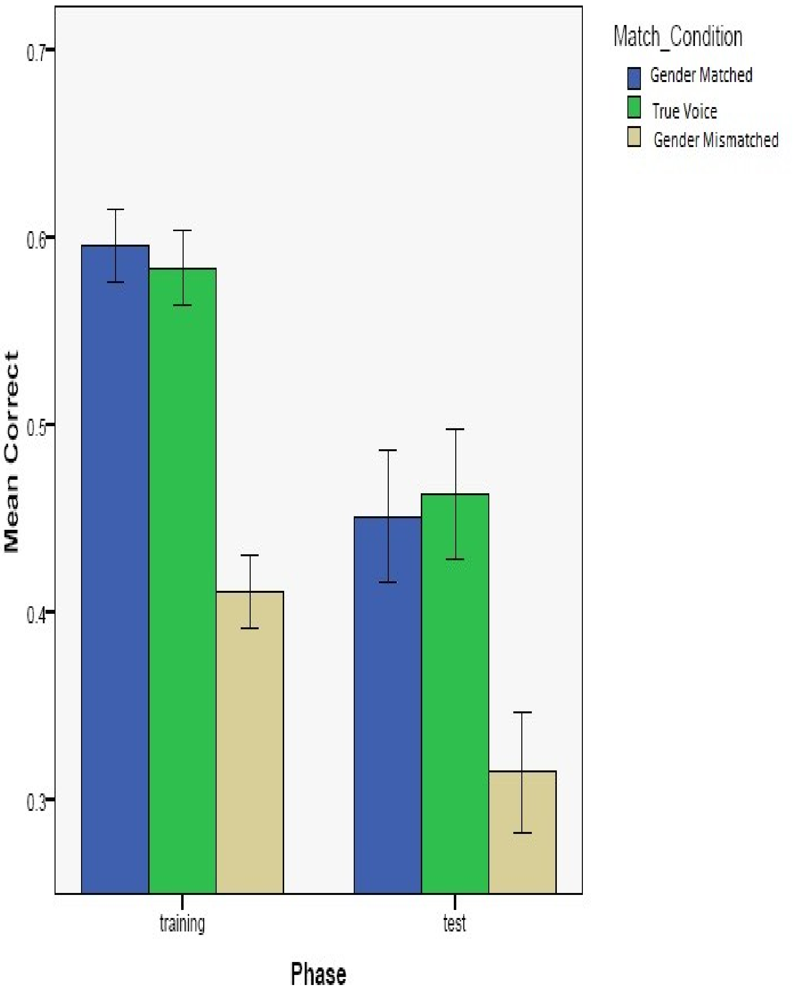


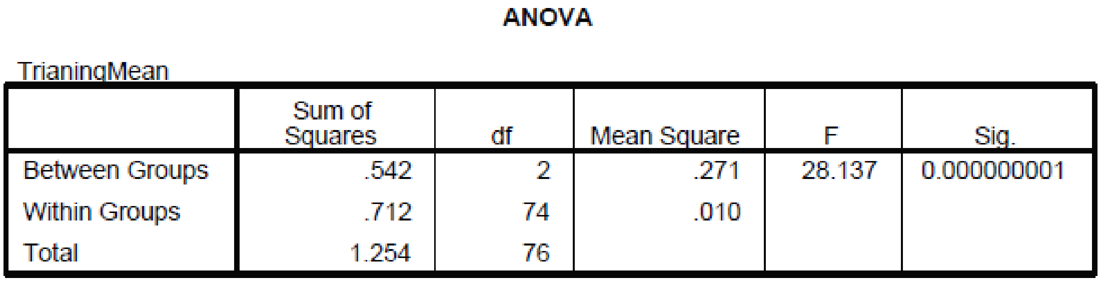
Figure 3. Difference between Conditions 1-3 in both training and test phase.

DISCUSSION

The current results support the hypothesis that learning to bind faces and voices is a specialized process that involves forming a supramodal ‘identity’. Participants were able to more quickly learn to pair the congruent faces and voices for specific picture/sentence pairs in the Test phase and reached much higher levels of performance overall. In the Test phase, where subjects had to generalize to new sentences based on learning the face-voice pairings, this difference was even more dramatic, with performance in the Gender Mismatched condition dropping to near chance levels. These results suggest that any learning that took place in the Gender Mismatched condition may have been due to memorization of the pairings considering the near chance levels of performance in the test phase.

The current experiment examined how people learn to pair faces and voices and incorporate this into an identity. The ability to be able to learn this pairing is a part of learning personal identity and therefore a process that is socially important. Previous studies have already shown that brief exposure to faces and voices induce more powerful multisensory associations than brief exposure to other multimodal combinations (Von Kriegstein & Giraud, 2006). Faces and voices may induce more powerful multisensory associations due the social importance, and perhaps people need to quickly associate these elements in order to categorize people into differing groups. Another aspect previous studies have examined is that people are able to more correctly match dynamic stimuli than static, perhaps follow-up studies to this experiment will find a difference in ultimate performance between the True Voice condition and the Gender Matched condition that was too sensitive to appear in the results with static stimuli. Based on previous studies as well as on this current experiment, several questions remain to be answered, such as whether learning to pair multimodal properties of an identity is more robust specifically for voices and faces or if this ability is similar with other properties of identity. Also, can people learn to pair properties of objects in a similar manner as they can learn to pair faces and voices; or, is this ability to pair multimodal properties of identity a human specific ability or do other animals possess this ability as well? Another aspect to examine is whether individuals with developmental delays or individuals on the autism spectrum are able to learn faces and voices similarly.

APPENDIX A

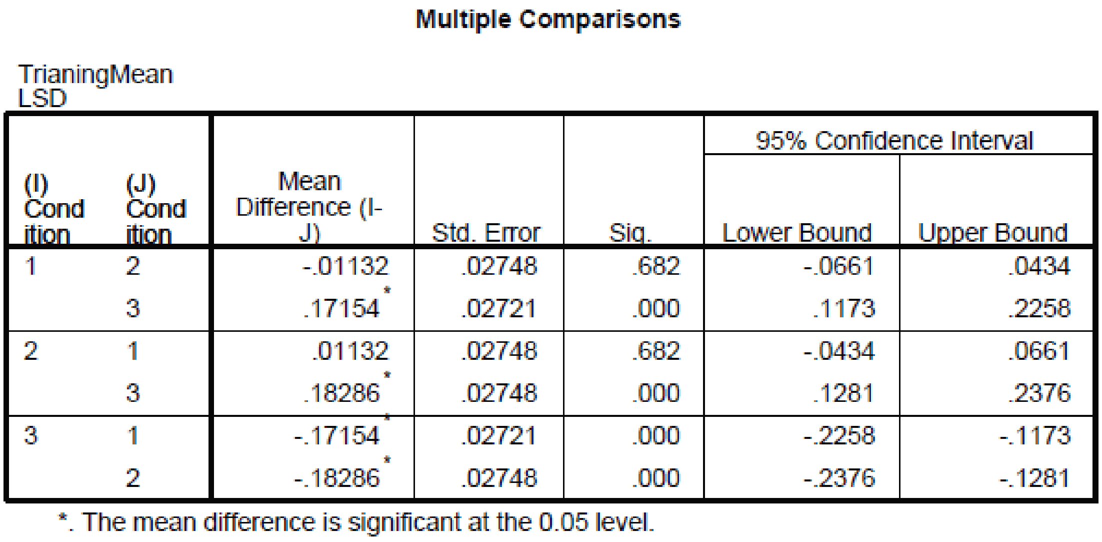


ANOVA

TrianingMean

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.542	2	.271	28.137	0.000000001
Within Groups	.712	74	.010		
Total	1.254	76			

Post Hoc Tests

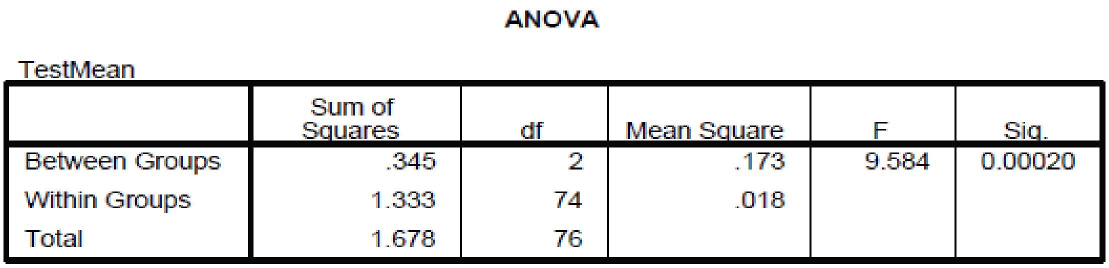


Multiple Comparisons

TrianingMean
LSD

					95% Confidence Interval	
(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	-.01132	.02748	.682	-.0661	.0434
	3	.17154*	.02721	.000	.1173	.2258
2	1	.01132	.02748	.682	-.0434	.0661
	3	.18286*	.02748	.000	.1281	.2376
3	1	-.17154*	.02721	.000	-.2258	-.1173
	2	-.18286*	.02748	.000	-.2376	-.1281

*. The mean difference is significant at the 0.05 level.

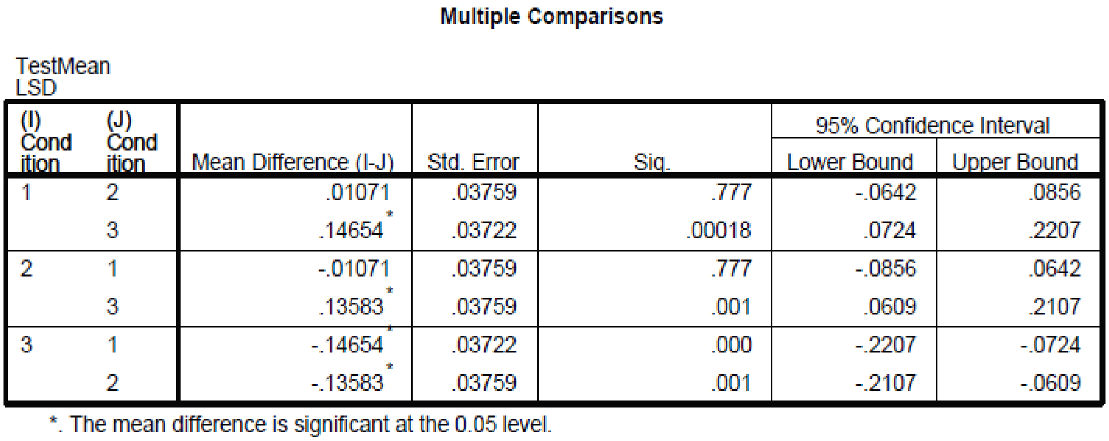


ANOVA

TestMean

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.345	2	.173	9.584	0.00020
Within Groups	1.333	74	.018		
Total	1.678	76			

Post Hoc Tests



Multiple Comparisons

TestMean
LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	.01071	.03759	.777	-.0642	.0856
	3	.14654*	.03722	.00018	.0724	.2207
2	1	-.01071	.03759	.777	-.0856	.0642
	3	.13583*	.03759	.001	.0609	.2107
3	1	-.14654*	.03722	.000	-.2207	-.0724
	2	-.13583*	.03759	.001	-.2107	-.0609

*. The mean difference is significant at the 0.05 level.

APPENDIX B

Face-Voice Stimulus Participation

Thanks for your help.
You will get a $5 Starbucks giftcard for your participation!

Please note that your name and contact information will NOT be used as part of the study. We only need them to contact you in order to give you your Starbucks credit.

Name_____________________________

How to contact you to get your credit____

- -

Subject # _________________ (to be completed by researcher)

Info for Study:
Age___________
Gender__________
Ethnicity/Race_____________

APPENDIX C

Visual/Audio Waiver Form

Florida Atlantic University, Dr. Elan Barenholtz and his associates have my permission to use my photograph, video and audio recordings taken in Dr. Barenholtz's laboratory for use as experimental stimuli, publication and presentation at scientific conferences and journals and for scientific web pages and other published materials. Neither my name nor any other identifying material will be associated with these images. I understand that there will be no further compensation to me for this use other than the gift card I received for my participation.

Signature Date

Print Name Permanent
Phone #
(optional)

Official Use Only

Lab Representative:

Signature Date

Print Name

APPENDIX D

Informed Consent Form: Visual Learning of Parts and Relations

1) Title of Research Study: Visual learning of Parts and Relations

2) Investigator: Dr. Elan Barenholtz

3) Purpose: The purpose of this research study is to study human learning of visual patterns

4) Procedures: Your participation in this study will consist of performing a computer-based task in which you will watch images of shapes and/or objects on the screen and then respond to what you saw using the keyboard. The experiment will take between 30 and 45 minutes.

5) Risks: The risks involved with participation in this study are no more than you would experience in regular use of a computer.

6) Benefits: Participation in this experiment will count as one study towards completion of your course requirement or extra credit (as described in your course materials). Additional potential benefits you may attain from participation in this research study include a greater knowledge of the methods used in cognitive psychology and contribution to the knowledge of the human mind and how it works.

7) Data Collection & Storage: Your data will be stored based on an assigned subject number not your personal identification. Your data will not be traceable back to your personal identification; it is effectively anonymous. All of your results will be kept confidential and secure and only the people working on this study will see your data unless required by law.

8) Contact Information: For related problems or questions regarding your rights as a subject, the Division of Research of Florida Atlantic University can be contacted at (561) 297-0777. For other questions about the study, you should call the principal investigator, Elan Barenholtz at (561) 297-3433.

9) Consent Statement: I have read or had read to me the preceding information describing this study. All my questions have been answered to my satisfaction. I am 18 years of age or older and freely consent to participate. I understand that I am free to withdraw from the study at any time without penalty. I have received a copy of this consent form.

Signature of Subject______________________________
Date________________

Signature of
Investigator___________________________Date______________

APPENDIX E

Face and Voice
DEBRIEFING FORM

About the visual learning of parts and relations experiment

Area of Psychology: Cognition, the science of how the mind works. This experiment was in the sub-area of human perception, how we see and learn to recognize objects in the world.

Experiment: The purpose of this study was to determine how people learn to identify people based on both their faces and their voices. First you learned to pair a face and a voice and then you were tested on your knowledge of that pairing.

Purpose: We can all recognize people based on both their faces and voices. The results of this experiment will help us to understand how we learn to do this.

Applications: Knowledge about how we see and recognize people can be applied to treat people with neurological disorders that prevent them from being able to recognize others.

Thank you for your participation!

REFERENCES

Belin, P. Fecteau, S. & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8(3),* 129-135.

Brookes, H., Slater, A., Quinn, P.C., Lewkowicz, D.J., Hayes, R., & Brown, E. (2001).Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant and Child Development, 1*, 75-82.

Calder, A.J. & Young, A.W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6,* 641-651.

Cloutier, J., Mason, M.F., & Macrae, C.N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology, 88,* 885-894.

DeBruine, L.M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblence. *Proceedings. Biological Sciences, 272,* 919-922.

Eberhardt, J.L. (2005). Imagining race. *American Psychologist, 60,* 181-190.

Gorno-Tempini, M.L., Price, C.J., Josephs, O., Vandenberghe, R., Cappa, S.F., Kapur, N., & Frackowiak, R.S.J. (1998). The neural systems sustaining face and proper-name processing. *Brain, 121,* 2103-2118.

Horstmann, G. & Ansorge, U. (2009). Visual search for facial expressions of emotions: A comparison of dynamic and static faces. *Emotion, 9,* 29-38.

Hughes, S.M., Dispenza, F., & Gallup, G.G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25,* 295-304.

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). ‘Putting the face to the voice’: Matching identity across modality. *Current Biology, 13,* 1709-1714.

Krauss, R.M., Freyberg, R., & Morsella, E. (2002). Inferring speakers’ physical attributes from their voices. *Journal of Experimental Social Psychology, 38,* 618-625.

Kreiman, J. (1997). Listening to voices: theory and practice in voice perception research. In K. Johnson and J. Mullinex (Eds.), *Talker Variability in Speech Research* (85-108). Academic Press.

Lachs, L. & Pisoni, D.B. (2004a). Crossmodal source identification in speech perception. *Ecological Psychology, 16(3),* 159-187.

Lachs, L. & Pisoni, D.B. (2004b). Specification of cross-modal source information in isolated kinematic displays of speech. *Journal of Acoustical Society of America, 116,* 507-518.

Macrae, C.N., Quinn, K.A., Mason, M.F., & Quadflieg, S. (2005). Understanding others: The face and person construal. *Journal of Personality and Social Psychology, 89,* 686-695.

Mesulam, M.M. (1998). From sensation to cognition. *Brain: A Journal of Neurology, 121(6),* 1013-1052.

Schweinberger, S.R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research, 40,* 453-463.

Todorov, A., Mandisodza, A.N., Goren, A., & Hall, C.C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623-1626.

Von Kriegstein, K. & Giraud, A.L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4,* 1809-1820.

Zebrowitz, L.A. & Collins, M.A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality & Social Psychology Review, 1,* 204-223.

Zuckerman, M., Miyaki, K., & Elkin, C.S. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research of Personality, 29,* 253-272.