EXPLORING THE UNIVERSALITY IN PERSONALITY IMPRESSIONS FROM BRIEF, NOVEL VOICES

BY

KATHERINE PEARL MARCOUX

MASTER'S THESIS

Advisor: Dr. Albert Costa
Supervisor: Dr. Cristina Baus
Msc in Brain and Cognition, Universitat Pompeu Fabra, Barcelona
July 2015

Abstract

When hearing a voice, we gather information about the individual, from physical information to their current mood. Additionally, from this interaction, we form personality judgments, which are consistent across listeners. Here, universality in personality judgments based on brief, novel voices is examined in two questions: Question 1 (Are the personality judgments and the dimensions of voice characteristics that people pay attention to the same across different languages?) and Question 2 (Do people make the same personality judgments and pay attention to the same voice characteristics in a foreign language as they do in their native language?). By testing personality impressions (aggressiveness, attractiveness, competence, confidence, dominance, likeability, trustworthiness, and warmth) from hearing "Hello" and "Hola" and comparing these results with McAleer, Todoroc, & Belin's study (2014) we find that the majority of results of ratings can be summarized in a two-dimensional social voice space defined by trustworthiness and dominance. This pattern suggests universality in personality judgments of voices. That is, it does not matter what your native language is nor what language the brief, novel voices are in because we form similar personality impressions. Additionally, the acoustical measurements the listener pays attention to for valence are explored.

Introduction

You walk into a bar, sit down and order a drink. You hear a female voice say "Hello". You have never met this woman before, have never heard her voice before. Yet, before you even turn your head to see who it is, you have already made a judgment about her personality. What information do you gather from her voice, a novel stimuli that lasted less than half a second? Is she trustworthy, attractive, dominant? These first impressions have been formed in your mind, helping you to formulate your decision to either start a conversation and get to know this person, or to simply be polite and continue on with your drink.

Facial Judgments

As human beings, we see faces and hear voices every day, making us experts in reading and extracting information from them. We can extract physical information such as age and gender from these faces (Burt & Perrett, 1997) and voices (Belin, 2004). Importantly, the information we receive from the stimuli is not limited to physical attributes, but we extract information which results in personality judgments.

We make judgments based on viewing faces and hearing voices and this in turn affects the way we treat individuals. Most of the research in this area relates to facial judgments. These facial judgments can affect various aspects of peoples' lives. As a matter of fact, faces have been found to impact the results of adjudication in U.S. small claims courts. For instance, Zebrowitz & McDonald (1991) found that defendants were more likely to lose their case as plaintiff's facial attractiveness increased. The effect of facial judgments has also been studied in US congressional elections, where inferences from faces on competence allowed for the prediction of the electoral outcome (Todorov et al., 2005). From these interactions, however short they may be, we gather considerable information that we then use to make inferences. These inferences occur very quickly and there is wide agreement among those making such inferences. Todorov, Pakrashi & Oosterhof (2009) confirmed and expanded upon previous research about judgments of faces, finding that after just 33ms, rater's agreements were above chance. Regarding these research findings, we can conclude that there is considerable agreement among individuals on first impressions from faces.

The literature on facial judgments of attractiveness and personality judgments has attempted to establish universality in such impressions. Perrett (1994) studied Caucasians and Japanese participants (i.e., raters) and found universality in facial judgments, while Zebrowitz et al. (2012) expanded upon cross-cultural studies of universality in impressions of faces by examining the Tsimane' people, a group who had no contact with modern culture. What they found was both within-cultural as well as between-cultural agreement of faces, although it should be noted that some effects were moderated by the judge and the face culture (Zebrowitz et al., 2012). This indicates that even though two individuals do not have the same social experience, their interpretation of faces is similar, although with slight variations, suggesting universality in the judgments.

In examining personality judgments on face evaluations, Todorov et al. (2008) found that a two-dimensional space in which valence/trustworthiness and power/dominance explained a majority of the data. This finding for two dimensional space for facial judgments parallels the valence and dominance findings for voices that McAleer, Todorov, & Belin, (2014) later established. The research indicates that there is universality in face evaluations, defined by a two-dimensional space of valence and dominance. As it has been found that face and voice impressions can parallel each other, it is important to explore the possibility of universality of personality judgments in voices with as defined by a similar two-dimensional space.

Vocal Judgments

As with faces, upon hearing voices, individuals gather many different types of social information. According to Belin, Fecteau, & Bédard (2004), we can consider voices "auditory faces" and subsequently, humans can identify the person and gather information about their emotional state. In addition to information relating to the meaning of speech itself (e.g. knowing that "Hello" is a greeting), from voices we gather biological characteristics of the individual such as age, gender, size, and fitness and affective information, dealing with the individual's emotions and motivational state (Belin 2004). Belin et al. (2011) later observed that "Although anatomically distinct, voice- and face-processing areas usually act in parallel and are assumed to communicate, facilitating our social responses" (p. 719). Rezlescu et al. (2015) examined the visual and auditory integration in personality impressions. They first examined the relation of faces and voices by having subjects rate voices and faces separately. They found that voice trustworthiness correlated with voice attractiveness, which paralleled the relationship between attractiveness and trustworthiness in faces. During the second part of their study, subjects saw faces and heard voices and the authors found that the visual and acoustic information each weighted differently in the judgment of personality. In terms of attractiveness, faces had more of an impact than voices, while for dominance the opposite was observed. Additionally, with trustworthiness, there was an interaction in which acoustic information was more important in determining trustworthy faces than for untrustworthy faces (Rezlescu et al., 2015). The judgments made by listening to voices were especially important for dominance and trustworthiness. There were some parallels between voice and face personality impressions, at times complementing each other, while, at other times their impact on personality judgments were not equal. What is relevant for our study is that trustworthiness and dominance evaluations seem made placing more emphasis on voices rather than faces.

Basis for the Present Study

Researching voices, McAleer, Todorov, & Belin (2014) discuss in their paper, "How do you say 'Hello'? Personality Impressions from Brief Novel Voices" that as with faces, voices rapidly generate first impressions about an individual's personality. Testing Scottish participants' judgments on personality traits (aggression, attractiveness, competence, confidence, dominance, femininity, likeability, masculinity, trustworthiness, and warmth) of brief, novel voices (as brief as the word "Hello") they found a two dimensional social space driven by valence (defined as trustworthiness and likeability) and dominance. That is, the personality judgments of voices that we make can be distilled down to two factors: valence (likeability and trustworthiness) and dominance.

Having established this two-dimensional space for personality impressions of voices, the aim of the present study is to expand upon this research by investigating whether this space holds across languages, making it part of a universal judgment. In doing so, we would expand the paralells with face judgments in showing that universality in first impressions holds for both faces and voices.

Acoustic Measurements and Personality Judgments

McAleer, Todorov, and Belin (2010) center their discussion of the two-dimensional space of valence and dominance around F0 (pitch) and its variations as measured by glide and intonation. For females, they explain that higher valence is related with a greater rise in pitch. While for males, they explain that having a higher pitch (F0), makes their voices more similar to women and in turn increases valence. This result is not surprising if we consider that in the literature, a heavy emphasis is placed on pitch and its role in emotions and personality impressions.

Indeed, previous research has extensively discussed voices in terms of F0 (pitch). Bauman & Belin (2010) found that to differentiate among speakers, pitch is a primary parameter, indicating its importance in voices and voice interpretation. It is a factor in reading emotions as well (Patel et al., 2011; Williams & Stevens, 1972). As for personality judgments, pitch plays a role in impressions of dominance (Ohala, 1982; Tigue et al., 2012), attractiveness (Puts, 2005; Feinberg et al., 2005; Tigue et al., 2012) and trustworthiness (Tigue et al., 2012). In turn, pitch appears to affect many decisions in our daily-life, such as mating choices, voting behavior, and even US Supreme Court outcomes (Puts, 2005; Tigue et al., 2012; Klofstad Anderson, & Peters, 2012; Chen, Halberstram & Yu, 2016).

Indeed, there is variability in acoustical measures in different languages. As Astruc et al. (2013) discuss in their paper English, Spanish, and Catalan are different in terms of rhythm, density of pitch accents, and syllable composition. In fact, English is a stress timed language while Spanish is syllable timed (Astruc et al., 2013). It is clear that variation in acoustical measures exist among languages, the question is whether we notice and pay attention to these differences if we are a native or foreign speaker.

Evaluation of Foreign Languages

One way to examine universality in voices is to compare different languages. However, language is a much more complex cognitive function than face processing and therefore exploring universality in language requires us to consider the language in which the decision is made, whether native or foreign. Preference for our native language is obvious in many contexts at various ages. Kinzler, Dupoux, & Spelke, (2007) found that infants preferred speakers who spoke in their native tongue and Kinzler, Shutts, & Spelke (2012) found the same with children. Indeed, this bias for native language affects our judgments. English speakers evaluated individuals speaking in their native language in terms of intelligibility and friendliness in more positive terms than those speaking in a foreign langauge (Lambert et al., 1960). Additionally, being presented with problems in various contexts in a foreign language changes the decisions you make (Costa et al., 2014; Keysar, Hayakawa, & An, 2012). Also, hearing foreign accents affects your judgments of the credibility of the statement due to processing difficulty (Lev-Ari & Keysar, 2010). While this may not directly reflect your judgment of the individual making the statements, it indicates that accents play a factor. Despite different judgments and preferences for one's native language, one should note that we are able to interpret emotions in foreign languages better than chance, although we are slightly better at doing so in our native language (Pell et al., 2009). Again, this suggests the universality in voice interpretation.

Questions of Interest

To examine different populations and languages to further explore the question of universality in impressions of voices, two questions were formulated: Are the dimensions of voice characteristics that people pay attention to the same across different languages? and Do people pay attention to the same voice characteristics in a foreign language as they do in their native language? In order to examine Questions 1 and 2, we used McAleer, Todorov, & Belin's (2014) study as a baseline, and tested Spanish native participants while hearing voices saying "Hola" (native language, Experiment 1) or "Hello" (foreign language, Experiment 2).

Question 1: Are the personality judgments and the dimensions of voice characteristics that people pay attention to the same across different languages?

In order to investigate Question 1 and address universality across languages in personality impressions from voices, we tested Spanish participants using "Hola" as the stimuli. This was chosen as the equivalent to the "Hello" tokens tested in Scotland with McAleer, Todorov, & Belin's study (2014) and will allow for cross-language comparisons. As there are differences in voice acoustics across languages, does this mean that the Spanish population makes the same personality judgments and pays attention to the same acoustical measurements as the Scottish or do they make different judgments and pay attention to different variables. One would expect that despite variations in acoustics, some acoustical measures such as pitch would be crucial, as it has been shown to affect many decisions in our lives (Puts, 2005; Tigue et al., 2012; Klofstad, Anderson, & Peters, 2012).

Question 2: Do people make the same personality judgments and pay attention to the same voice characteristics in a foreign language as they do in their native language?

As for addressing Question 2 and examining the universality of judgments of the same stimuli in one's native versus foreign language, we tested Spanish participants with the "Hello" tokens from the original study (McAleer, Todorov, & Belin, 2014) and were able to compare their results with ours. Do Spaniards make the same personality judgments? If so, do they pay attention to the same acoustical measurements as the Scottish population in order to make these cross-cultural judgments? Are the properties of own language applied to a foreign language?

Experiment 1: Hola-Bcn

Methods

Participants

100 participants (21.39 ± 4.01 years, 39 males) at the Universitat Pompeu Fabra took part in Experiment 1. The selection criteria for all participants was that they were native Spanish speakers who were born and raised in Spain.

Material

64 native Spanish speakers (22.1±4.2 years, 32 males) from the Universitat Pompeu Fabra were used to record the stimuli for the experiment. The selection criteria for the speakers was that they were born and raised in Spain, so as to avoid non-native accents and to limit the variability in native accent to only Spain. Written consent was obtained from each participant.

The majority of the 64 speakers for Experiment 1 were individually recorded with a handheld stereo speaker at 44.1k Hz while the remainder were recorded in a soundproof room using Audacity v2.12. The speakers were requested to read a telephone conversation in Spanish in a neutral tone, which can be found in Appendix A. From the recordings, the noise was reduced and subsequently the Hola tokens were extracted using Audacity® v2.12. The stimuli were also normalized for power and loudness using PRAAT 6.0.14 and lasted an average of 319 ms ± 67.1 ms for male voices and 338 ms ± 60.0 ms for female voices.

"Hola" was selected as the Spanish equivalent of "Hello", the term used in the original study (McAleer, Todorov, & Belin, 2014). This allows for first impressions from a brief token, which is a social word, familiar to all.

The following eight acoustic measurements were extracted from each "Hola" token and measured using PRAAT 6.0.14.

1. Mean fundamental frequency (F0), which is the measurement of pitch, the periodic oscillation at fundamental frequency of the vocal folds (Latinus & Belin, 2011).

2. Intonation, calculated by subtracting F0-min from F0-max.

3. Glide, calculated by subtracting F0-start from F0-end.

4. Formant dispersion, which is the ratio between consecutive formant means. The equation used to calculate the formant dispersion can be found in Appendix B.

5. Harmonic-to-noise ratio (HNR), which indicates roughness in the voice.

6. Jitter, which measures local variations in frequency (Bauman & Belin, 2010).

7. Shimmer, which measures local variation in amplitude (Patel et al., 2011).

8. Alpha ratio, which measures the slope of the source spectral (McAleer, Todorov, & Belin, 2014).

Procedure

Each participant was assigned one trait with which to rate the 64 Hola tokens. Thus, an average of 12.5 participants evaluated a given personality trait. The eight personality traits chosen from the original experiment were: aggressiveness, attractiveness, competence, confidence, dominance, likeability, masculinity, trustworthiness, and warmth. Each participant rated only one of the eight personality traits, so as to avoid the Halo effect. The potential danger that we wanted to avoid with the Halo effect was that if someone rated a voice high on trustworthiness, then they would most likely rate that same voice high on likeability as well (McAleer, Todorov, & Belin, 2014).

When asked to rate each stimuli, participants were given the following instructions: "Evalúa cuán {TRAIT} te parece esta persona, según su voz". The scale subjects were presented with was a 9-point Likert scale, ranging from 1, extremadamente no{TRAIT} to 9, extremadamente {TRAIT}. The male and female voices were blocked separately. During each gender block, the Hola tokens were presented in a randomized order and each voice was repeated once. The blocking of stimuli by gender was counterbalanced across participants. Participants were given a break between the presentation of the male and female voices, the length of which was left to their discretion. The experiment was performed using ePrime v2.0.

Exclusion Criteria

Participants were excluded if they were born or raised outside of Spain. For Experiment 1, 1 participant was excluded for this reason. Additionally, participants were asked if they recognized any of voices, and if so, they were excluded from the analyses, in this case 16 participants were excluded. If the participants had already completed the study for a different trait for Experiment 1 or Experiment 2, they were excluded, this applied to 3 participants.

The two exclusion criteria from the original study, were also applied: (1) If a participant rated more than two-thirds of the same stimuli greater than 2 ratings apart, they were excluded. For example, if for over two-thirds of the time, the participant rated the voice as 3 and for the repetition of the voice as 6, then the participant would be excluded. 6 participants were excluded based on this criterion. (2) If a participant rated more than 75% of the voices with the same rating, they were excluded. This would be the case, for example, if the participant rated all voices 7. No participants were excluded based on criterion 2. These two exclusion criteria were established prior to the experiment and helped to confirm that participants were being attentive.

Data Analysis

Differences in vocal acoustics between male and female voices result in differences in F0 as well as formant distribution. Therefore, the analyses performed were separated by the gender of the voices (Puts et al., 2007). The original study conducted by McAleer, Todorov & Belin (2014) also separated the data by the gender of the voices before conducting the analyses.

Principal Component Analyses (PCA) were performed to determine the relationships of the ratings of the personality traits. Principal components reveal the subsets that reflect underlying processes (Bauman & Belin, 2010). This allowed the personality traits that form these subsets to be determined, and in turn understand what drove the two main components, which explain a majority of the data.

Once the data was separated by gender, the z-score for each individual participant was calculated. Then, the average ratings for each voice were determined for each personality trait.

The acoustic measurements of the vocal stimuli were examined using a multiple linear regression model to determine which of the acoustic measurements (F0, intonation, glide, formant dispersion, HNR, jitter, shimmer, and alpha) provided the highest impact on the first principal component. For the analyses of Experiments 1 and 2 it was decided to focus solely on the first principal component as that explained most of the variation in the ratings (explaining between 52% and 69% of the data depending on the gender of the voices and the experiment).

Results

In order to determine inter-rater reliability, Chronbach's Alpha was calculated for each personality trait. Values for alpha above 0.85 are considered high for inter-rater agreement, which was the case for all personality traits except dominance ($r=0.68$). The table with the alpha values can be found in Appendix C.

Male Voices PCA

For the ratings of the male "Hola" tokens, a two dimensional solution was found, in which 85.12% of the variance was explained; principal component 1 (PC1) explained 65.54% of the data and PC2 explained 19.58%[1]. The two personality traits that most strongly drove PC1 were confidence, (Pearson's r value of 0.92) and trustworthiness ($r=0.92$), while aggressiveness ($r=0.77$)

[1] In the original study, PC1 for male voices explained 56.18% of the variance and PC2 explained 31.18%, for a total of 87.98%.

and dominance ($r=0.68$) were the main traits driving PC2. The tables with the loadings of the PCA correlated with each personality trait for Experiment 1 can be found in Appendix D. In *Figure 1* below, you can see the two dimensional solution for PCA for "Hola" male tokens.

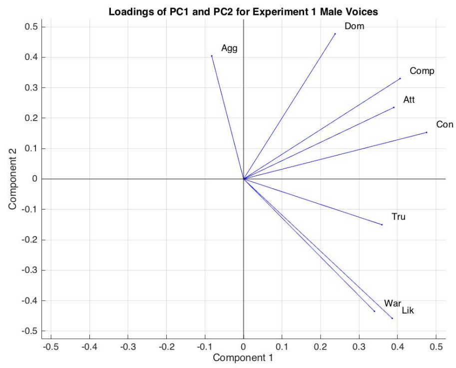

Figure 1

Female Voices PCA

For the ratings of the female "Hola" tokens, a two dimensional solution was also found, in which 83.15% of the variance was explained; PC1 explained 52.60% of the variance and PC2 explained 30.55%[2]. The two personality traits that most strongly drove PC1 were warmth ($r=0.97$) and likeability ($r=0.95$), while competence ($r=0.91$) and dominance ($r=0.79$) were the main traits driving PC2. In *Figure 2* below, you can see the two dimensional solution for PCA for "Hola" female tokens.

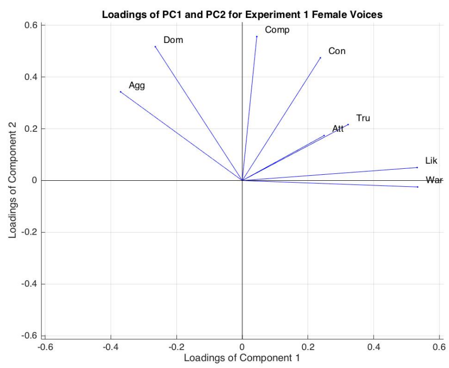

Figure 2

In sum, in Experiment 1 it was found that valence, as indexed by trust in males and likeability for females, was driving PC1. In both cases, dominance was one of the two main traits driving PC2.

[2] In the original study, PC1 for female voices explained 59.54% of the variance and PC2 explained 25.53%, for a total of 85.07%.

Multiple Linear Regression

Separated by gender, the eight acoustic measurements were used in a multiple linear regression model to explain which ones most impacted the first principal component. For male voices, the only acoustic measure that significantly impacted PC1 was alpha ($b = -1.72$, $p < 0.01$), while for female voices the significant acoustic measure was F0 ($b = 0.029$, $p < 0.01$). All the values from the multiple linear regression model can be found in Appendix E.

Variation in the Acoustical Measurements

A Multivariate ANOVA was performed with language (Spanish and English) and gender (male and female voices) as fixed factors and the eight acoustic measurements as the dependent variables. The aim of performing the multivariate ANOVA was to determine if there were indeed variations in the acoustics measurements between Spanish and English. This allowed us to explore the similarities and differences across the two languages used to answer Question 1. Across the languages, glide, dispersion, HNR, jitter, and shimmer were found to be significant different, all with $p < 0.001$. An interaction effect of language by gender was found for intonation ($p = 0.037$), glide ($p < 0.001$), and jitter ($p = 0.025$).

Experiment 2: Hello-Bcn

Methods

Participants

305 participants (22.03 ± 4.5 years, 129 males) from the same pool as Experiment 1 were used.

Stimuli

As Experiment 2 is a replication of the original study using Spanish participants, the stimuli and procedure are exactly the same as in the original study and Experiment 1. Therefore, with the permission of the authors, the "Hello" tokens from the original data were used. The stimuli consisted of 64 Scottish speakers (28.2 ± 10.2 years, 32 male) born and raised in Scotland. For the preparation of their data, they individually digitally recorded the participants using a 16 bit mono, 44100Hz, WAV format in a soundproof room. The participants were asked to read a telephone passage using direct speech, in a neutral tone. In their case, they normalized for power and loudness using Matlab (the Mathworks). For male voices, the stimuli lasted an average of 391ms ± 65.1ms and for female voices it lasted 390 ms ± 64.1ms.

Additionally, the same eight voice characteristics (1. mean F0, 2. intonation, 3. glide, 4. formant dispersion, 5. harmonic-to-noise, 6. jitter, 7. shimmer, and 8. alpha) were used in the analyses.

Procedure

The exact same procedure was used as in the original study and Experiment 1, with the following instructions presented: "Based on the voice, please rate how {TRAIT} is this person". The scale subjects were presented with was a 9-point Likert scale, ranging from 1, extremely un{TRAIT} to 9, extremely {TRAIT}.

Exclusion Criteria

10 participants were excluded as they were not born or raised outside of Spain. If the participants had already completed the study for a different trait for Experiment 1 or Experiment 2, they were excluded, this applied to 8 participants. For criterion 1 from the original study 33 participants were excluded from the analyses and 2 were excluded for criterion 2.

Data Analysis

The same analyses as in Experiment 1 were performed.

Results

Inter-rater reliability as measured by Chronbach's Alpha was high for all personality traits in Experiment 2. The table with values of for each personality trait can be found in Appendix C.

Male Voices PCA

For the ratings of the "Hello" male token, a two dimensional solution was found, in which 90.45% of the variance was explained; PC1 explained 65.42% of the variance and PC2 explained 25.03%[3]. The two personality traits that most strongly drove PC1 were competence ($r=0.93$) and trustworthiness ($r=0.92$), while aggressiveness ($r=0.86$) and dominance ($r=0.72$) were the main traits driving PC2. In *Figure 3* below, you can see the two dimensional solution for PCA for Hello tokens for Experiment 2.

---

[3] In the original study, PC1 for male voices explained 56.18% of the variance and PC2 explained 31.18%, for a total of 87.98%..

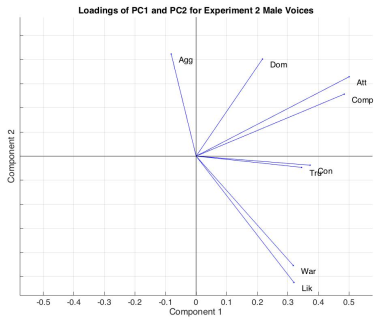

Figure 3

Female Voices PCA

For the ratings of the female "Hello" tokens, a two dimensional solution was also found, in which 86.46% of the variance was explained; PC1 explained 69.15% of the variance and PC2 explained 17.31%[4]. The two personality traits that most strongly drove PC1 were attractiveness (*r=0.94*) and likeability (*r=0.93*), while aggressiveness (*r=0.86*) and dominance (*r=0.85*) were the main traits driving PC2. In *Figure 4* below, you can see the two dimensional solution for PCA for Hello tokens for Experiment 2.

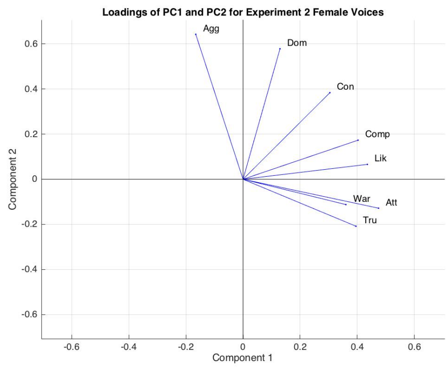

Figure 4

In Experiment 2 for both male and female voices, it was found that one trait from valence, trust and likeability respectively was driving PC1. As is the case for both female and male voices, aggression was the strongest driving factor for PC2, followed by dominance.

Multiple Linear Regression

In performing a multiple linear regression to explain which acoustic measurements most impacted PC1 it was found that there are no acoustic measures significantly impacting PC1 in male tokens, although F0 approaches significance (*b=0.012, p=0.06*) as does jitter (*b=6526, p=0.09*). For female

[4] In the original study, PC1 for female voices explained 59.54% of the variance and PC2 explained 25.53%, for a total of 85.07%.

tokens, glide (*b=-0.0066, p =0.01*) and intonation (*b=0.0083, p<0.01*) significantly impacted PC1, while alpha (*b=-1.86, p=0.5*) and F0 (*b=-0.019, p=0.06*) approached significance.

Summary of Experiments 1 and 2

Across genders for experiments 1 and 2, we found that one trait from valence (either trustworthiness or likeability) was one of the two main traits driving PC1 while dominance was one of the two main traits driving PC2. As for the multiple linear regression performed for Experiment 1 for male voices alpha was significant and for females voices F0 was significant. For Experiment 2, for male voices F0 was marginally significant and for female voices intonation and glide were significant.

Questions 1 and 2

Considering Questions 1 and 2, it is important to be cognizant of the fact that Experiments 1 and 2 are related to the original study by McAleer, Todorov, & Belin (2014). To establish that there is indeed a relationship, a correlation of the averages for each trait in the original study with Experiments 1 and 2 was performed.

In examining Question 1 (original versus Experiment 1) for male voices the correlation was *r=0.87* while for female voices it was *r=0.90*. Values of r above *0.76* are considered significant at the level of *p<0.001*, as is the case here (Fischer, 1925).

When examining Question 2 (original versus Experiment 2) the data for male voices was found to have an r-value of *0.78* and for female voices r was 0.73, significant at the *p<0.001* level for males and at the *p<0.05* level for females (Fischer, 1925).

Questions 1 and 2: PCA

In order to be able to compare Experiment 1 and Experiment 2 with the original study, a combined PCA was performed with each to examine how the different personality traits grouped and drove PC1 and PC2 and how this may have changed. The results can be found in Appendix E.

The main finding was that when combined, for Question 1 for male tokens, trustworthiness (*r=0.90*) and confidence (*r=0.87*) were the main traits driving PC1 and dominance (*r=0.82*) and aggression (*r=0.79*) were the main traits driving PC2. For Question 1, PC1 is driven by trustworthiness (*r=0.93*) and likeability (*r=0.93*) and PC2 by dominance (*r=0.95*) and aggression (*r=0.73*).

As for Question 2 for male tokens, the exact same traits were driving PC1 (trustworthiness (*r=0.89*) and confidence (*r=0.85*) and PC2 dominance (*r=0.84*) and aggressiveness (*r=0.80*)). For Question 2 for female tokens, PC1 is driven by trustworthiness (*r=0.94*) and likeability (*r=0.93*) and PC2 by (*r=0.83*) and aggressiveness (*r=0.80*).

The general trend is that for Questions 1 and 2 trustworthiness is driving PC1 and dominance is driving PC2.

Discussion

The study aimed to not only examine universality across languages but also in terms of native and foreign language evaluations in terms of personality judgments and determining which acoustic measurements individuals pay attention to. By comparing Experiments 1 and 2 with the original study (McAleer, Todorov, & Belin, 2014), progress can be made with the research in personality impressions in voices, specifically from novel, brief stimuli. The first thing we established which allowed us to compare the studies to answer Question 1 (Are the personality judgments and the dimensions of voice characteristics that people pay attention to the same across different languages?) and Question 2 (Do people make the same personality judgments and pay attention to the same voice characteristics in a foreign language as they do in their native language?) was a high inter-rater reliability as determined by Chronbach's Alpha (with the exception of dominance in Experiment 1). This meant that participants were in agreement in their judgments of personality of the voices.

Universality in Social Voice Space

Once it was established that there was agreement in judgments, this allowed us to examine how the ratings of personality traits were expressed in a two-dimensional social voice space. With Scottish participants, this space was defined by valence (trustworthiness and likeability) and dominance (McAleer, Todorov, & Belin, 2014). To attempt to answer the question of universality in personality impressions of voices, Experiments 1 and 2 were compiled with the original study for PCA calculations. We found a consistent trend across both male and female voices for Questions 1 and 2 in which trustworthiness was always one of two main traits driving PC1 and dominance was always one of two main traits driving PC2. This parallels the two-dimensional social voice space that McAleer, Todorov, & Belin (2014) established as valence (trustworthiness and likeability) driving PC1 and dominance as driving PC2. This parallels the original study for valence and dominance and answers both our questions. Not only is there universality across judgments of voices in different languages (Question 1), but this universality is also found in making judgments about personality in a foreign language (Question 2).

In general, this two-dimensional social voice space found by McAleer, Todorov, & Belin (2014) and in the current study reveals important information for universality in voice impressions. This social voice space hints at deeper universality of categories. Todorov et al. (2008) found a similar two-dimensional space defined by valence/trustworthiness and power/dominance for faces and Wiggins (1979) found a similar two-dimensional solution for interpersonal relations.

As discussed, our findings suggest that there are universal judgments in personality impressions of voices.

Universality in Acoustic Characteristics in Determining Personality Judgments

The next step was to determine which acoustic measurements most impacted these ratings and to see if they also suggested that there may be universality in the measurements we focus on when judging voices.

Examining Question 1 in terms of acoustical measurements that contribute to the two-dimensional space is more complicated than Question 2 because the stimuli are different, but maybe more interesting (comparing "Hello" and "Hola" tokens). The acoustical measurements for the different languages that were found to be significantly different were glide, dispersion, HNR, jitter, and shimmer. It is crucial that F0 (pitch) was found to not be considered significantly different between the two stimuli sets as F0 is one of the key acoustical measurements that individuals focus on when making personality judgments (Feinberg et al., 2005; Ohala, 1982; Puts, 2005; Tigue et al., 2012).

Question 1: Are the dimensions of voice characteristics that people pay attention to the same across different languages?

In the original study, they found that for male voices F0 and HNR were significant in contributing to valence in the two-dimensional space established, while for females HNR, glide, and intonation were the contributing measures. For Experiment 1, for male voices alpha contributed significantly and for female voices F0 was significant. As we can observe, while the measurements that the subjects paid attention to did not perfectly match up, there are some connections. The acoustical measurements that contributed to PC1 of male voices in Experiment 1 do not match the original study, however, Experiment 1 for female voices, F0 is significantly contributing to PC1, which parallels the significant F0 in male voices in the original study. There is no clear explanation as to why the acoustic measurements significant for PC1 for male voices do not parallel the original study and the acoustic measurements significant for PC1 for female voices have the same acoustic measurements as the male voices in the original study. It could be that the differences in acoustical measurements in the "Hello" and "Hola" tokens are contributing to this difference.

It should be taken into consideration that for this study, the voices used were not manipulated to explore full variability. Rather, as was the case in the original study (McAleer, Todorov, & Belin, 2014), the voices were collected from the desired population and then the acoustics were measured. As we did not control for the variability of the voices, we can examine which acoustics most significantly impacted the two-dimensional space, however it is clear that further research needs to be done. Conducting further research in which vocal acoustics are manipulated and accounted for would better identify which acoustical measurements we focus on when listening to voices and form personality impressions.

Conducting experiments controlling the variability of the acoustical measures while measuring personality impressions would also bring us one step closer in determining which acoustical measures impact the different personality traits and how we may create an ideal attractive or trustworthy voice.

Question 2: Do people pay attention to the same voice characteristics in a foreign language as they do in their native language?

It is more difficult to answer Question 1 in terms of which acoustical measurements we focus on while making personality judgments of voices due to differences in acoustical measures. Question 2 allows for a clearer picture, as the stimuli, and hence the acoustical measurements, were exactly the same. Indeed, in judgments of male voices in the original study (McAleer, Todorov, & Belin, 2014) F0 and HNR were the significant factors contributing to PC1, as defined by valence. For Experiment 2 for male voices there were no measurements that were significant, however F0 and jitter approached significance, indicating parallels between the original study and Experiment 2, with F0 being the significant acoustic measurement in common. For female voices, parallels between the original study and Experiment 2 were also found. McAleer, Todorov, & Belin, (2014) found that HNR, glide and intonation significantly contributed to PC1 and for Experiment 2, intonation and glide, were also found to be significant for PC1. In their discussion of how the acoustical measures explain valence and dominance, they focus on the explanation of pitch (F0). Additionally, in essence, glide (as calculated by F0-end minus F0-start) and intonation (as calculated by F0-max minus F0-min) are measurements in the variation of pitch and hence relate to pitch.

One way to complete Question 2, addressing universality in impressions of voices in your native and foreign language would be to have Scottish participants evaluate the "Hola" tokens. This would allow for more comparisons and to confirm our findings.

Question 1 and 2 Summary

We examined Questions 1 and 2 in terms of what acoustic characteristics we pay attention to for PC1. We find that it is hard to ascertain much from Question 1, examining universality across languages, as Experiment 1 does not necessarily parallel the findings in the original study. This may be explained because of the differences in acoustical measurements in "Hello" and "Hola" tokens. For Question 2, examining universality in your native and foreign language, we find a much clearer picture, in which there are parallels for male and female voices between the original study and Experiment 2.

Due to the inconclusive results from Question 1, we can say that perhaps because of the variability in acoustical measurements, individuals take different acoustical measurements into consideration. However, it should be noted that examining Question 1, in terms of personality judgments across languages, there was universality. Since different acoustic measurements were significant for PC1, in the original study as compared to Experiment 1, this means that the Spaniards and Scottish form similar personality impressions but are paying attention to different acoustic measurements.

Question 2, in turn, suggests that there may be universality in the way we interpret voices as we do with faces (Perrett, 1994; Zebrowitz et al., 2012) and that the acoustic measurements that we pay attention to are similar. The acoustical characteristics that were significant for the Scottish participants were the same for Spaniards, despite English not being their native language.

Limitations and Further Research

In order to improve the current study and expand upon the analysis, it would be instructive to perform analyses based on the gender of the rater, especially for a trait such as attractiveness. This separation of raters by gender may not significantly change the results, but may alter the picture slightly and provide additional insight into how male and female raters may be driving the results differently. Research has shown that there is an overall preference for pitch in male voices (Klofstad, Anderson, & Perters, 2012; Tigue et al., 2102; Tsantani et al., 2016). Jones et al. (2010) expand upon this and found a domain-specific opposite sex bias, in which males strongly preferred feminized females voices (in which the pitch was higher), while women did not show a preference for raised or lowered pitch in other women's voices. However, for other traits such as dominance, it may not be the case that raters need to be separated by gender, as Jones et al. (2010) also found that there was no opposite sex bias when considering dominance. Lowered pitch in male voices were perceived as more dominant in both male and female subjects, and the same was found with lowering of women's voices. Hence it would be informative to explore some of the traits as separated by the gender of the rater.

Conclusion

Testing Spaniards on personality judgments of "Hello" and "Hola" and comparing these findings to the study conducted by McAleer, Todorov, & Belin (2014) allowed us to gain insight into universality of voice impressions. Question 1 (Are the personality judgments and the dimensions of voice characteristics that people pay attention to the same across different languages?) and Question 2 (Do people make the same personality judgments and pay attention to the same voice characteristics in a foreign language as they do in their native language?), were the driving questions in this study. We found that there was indeed a two-dimensional social voice space as defined by trustworthiness and dominance that described a majority of the ratings for judgments of personality traits (aggression, attractiveness, competence, confidence, dominance, likeability, trustworthiness, and warmth), which parallels the valence and dominance two-dimensional space found by McAleer, Todorov, & Belin (2014). The results and interpretation of which acoustical measures subjects paid attention to was not as clear for Question 1 as there were differences in acoustical measures that were not manipulated for variability. However, in examining Question 2, there is a clear parallel between Experiment 2 and the original study, which hints at a universality in which acoustic measures are significant.

There is clear universality in the judgments of personality as defined in a two-dimensional space of a valence-related trait (trustworthiness) and dominance. Future research needs to be conducted in manipulating the variability of the acoustics for Experiment 1. Based on the parallel results of Experiment 2 with the original study, there is an indication that there may be universality for voices. However, based on the results of Experiment 1, the conclusion cannot be as solid as when the results of both Experiments have parallel results with the original study.

References

Astruc, L., Payne, E., Post, B., del Mar Vanrell, M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech, 56*(2), 229-253.

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research PRPF, 74*(1), 110-120.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711-725.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129-135. doi:http://dx.doi.org/10.1016/j.tics.2004.01.008

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., . . . Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology, 20*(2), 116-120. doi:http://dx.doi.org/10.1016/j.cub.2009.11.034

Chen, D. L., Halberstram, Y., & Yu, A. (2016). "Mr. chief justice, may it please the court?" First impressions of lawyers' voices predict US supreme court outcomes.

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition, 130*(2), 236-254.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561-568.

Fisher, R.A. (1925). Classics in the History of Psychology. Retrieved July 11, 2016, from http://psychclassics.yorku.ca/Fisher/Methods/chap6.htm

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*(2), 1213-1222.

Jastrow, J. (1885). Composite portraiture. *Science, ,* 165-167.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57-62.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*(6), 661-668. doi:10.1177/0956797611432178

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577-12580.

Kinzler, K. D., Shutts, K., & Spelke, E. S. (2012). Language-based social preferences among children in South Africa. *Language learning and development*, *8*(3), 215-232.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20120311.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*(1), 44.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115-121.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*(4), R143-R145. doi:http://dx.doi.org/10.1016/j.cub.2010.12.033

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093-1096.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? personality impressions from brief novel voices. *PloS One, 9*(3), e90779.

Ohala, J. J. (1982). The voice of dominance. *The Journal of the Acoustical Society of America*, *72*(S1), S66-S66.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*(1), 93-98.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120.

Perrett, D. (1994). Facial shape and judgements. *Nature, 368*, 17.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*(5), 388-397.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340-344.

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior, 39*(4), 355-370.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33*(3), 210-216.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813-833.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455-460.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science (New York, N.Y.), 308*(5728), 1623-1626. doi:308/5728/1623 [pii]

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception*, 0301006616643675.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of personality and social psychology*, *37*(3), 395.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America, 52*(4B), 1238-1250.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*(6), 603.

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among US and culturally isolated Tsimane'people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology, 43*(1), 119-134.

Appendix

Appendix A: Telephone conversation

Conversation 1: Read by male subjects:
A: Hola! Soy Pablo. ¿Quien habla?
B: Hola Pablo, habla Pedro.
A: Ah, hola. ¿Está Juan?
B: Si, ahora te paso.
C: Hola Pablo.
A: Hola Juan, ¿como estás?

Conversation 2: Read by female subjects:
A: Hola! Soy María. Quien habla?
B: Hola María, habla Silvia.
A: Ah, hola. ¿Está Isabel?
B: Si, ahora te paso.
C: Hola María.
A: Hola Silvia, ¿como estás?

Appendix B: Formant Dispersion Equation

Equation from Fitch (1997) used to calculate formant dispersion:

$$D_j = \frac{\sum_{i=1}^{n}(F_{i+1} - F_i)}{n - 1}$$

Appendix C: Alpha Values

Experiment 1	**Chronbach's Alpha**	**Subjects**
Aggressiveness	0.98	10
Attractiveness	0.92	8
Competence	0.98	10
Confidence	0.93	9
Dominance	0.68	9
Likeability	0.92	10
Trustworthiness	0.96	10
Warmth	0.95	10
Average	0.91	9.5

Experiment 2	Chronbach's Alpha	Subjects
Aggressiveness	0.97	33
Attractiveness	0.97	30
Competence	0.96	37
Confidence	0.92	32
Dominance	0.93	32
Likeability	0.90	32
Trustworthiness	0.93	29
Warmth	0.97	33
Average	0.94	32.25

Appendix D: Experiments 1 and 2: PCA Correlations for each Personality Trait

Experiment 1: Male Voices	Component 1	Component 2
Aggressiveness	-0.2887	**0.7721**
Attractiveness	0.7908	0.2606
Competence	0.8721	0.3870
Confidence	**0.9171**	0.1614
Dominance	0.6196	**0.6794**
Likeability	0.8121	-0.5267
Trustworthiness	**0.9161**	-0.2094
Warmth	0.7836	-0.5481
Variance	0.65537	0.19579

Experiment 1: Female Voices	Component 1	Component 2
Aggressiveness	-0.7826	0.5519
Attractiveness	0.5447	0.2899
Competence	0.0954	**0.9142**
Confidence	0.4949	0.7521
Dominance	-0.5316	**0.7920**
Likeability	**0.9473**	0.0681
Trustworthiness	0.8227	0.4208
Warmth	**0.9669**	-0.0343
Variance	0.5260	0.3055

Experiment 2: Male Voices	Component 1	Component 2
Aggressiveness	-0.2672	**0.8584**
Attractiveness	0.8980	0.3653
Competence	**0.9329**	0.3062
Confidence	0.8947	-0.0558
Dominance	0.6235	**0.7154**
Likeability	0.6782	-0.6872
Trustworthiness	**0.9154**	-0.0766
Warmth	0.7173	-0.6336
Variance	0.65418	0.25025

Experiment 2: Female Voices	Component 1	Component 2
Aggressiveness	-0.4418	**0.8569**
Attractiveness	**0.9365**	-0.1265
Competence	0.9185	0.1975
Confidence	0.7346	0.4632
Dominance	0.3803	**0.8491**
Likeability	**0.9300**	0.0704
Trustworthiness	0.9298	-0.2447
Warmth	0.8767	-0.1366
Variance	0.6915	0.1731

Appendix E: Multiple Linear Regression Model

Experiment 1: Male Voices PC1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.50E+00	3.23E+00	0.774	0.4469
F0	-5.36E-04	9.95E-03	-0.054	0.95753
intonation	-2.57E-03	3.82E-03	-0.673	0.50781
dispersion	-1.55E-04	2.89E-03	-0.054	0.95771
glide	6.36E-03	3.97E-03	1.605	0.12224
HNR	3.13E-02	1.48E-01	0.211	0.83459
alpha	-1.73E+00	5.90E-01	-2.928	0.00756
jitter	1.63E+03	2.84E+03	0.573	0.57221
Shimmer	4.20E-01	1.03E+00	0.407	0.68812

Experiment 1: Male Voices PC2

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept)	1.09E+00	1.46E+00	0.745	0.463687
F0	-5.57E-03	4.49E-03	-1.241	0.227015
intonation	-5.63E-03	1.72E-03	-3.267	0.003391
dispersion	-9.06E-04	1.30E-03	-0.695	0.493909
glide	7.64E-03	1.79E-03	4.267	0.000289
HNR	-2.81E-02	6.69E-02	-0.421	0.677806
alpha	2.37E-01	2.66E-01	0.89	0.382484
jitter	3.38E+03	1.28E+03	2.639	0.014664
Shimmer	9.04E-02	4.66E-01	0.194	0.847691

Experiment 1: Female Voices PC1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.42E+00	3.49E+00	-1.266	0.21832
F0	2.89E-02	8.17E-03	3.536	0.00177
intonation	6.98E-04	2.79E-03	0.25	0.80443
dispersion	-2.59E-03	3.07E-03	-0.844	0.40719
glide	-1.45E-03	2.83E-03	-0.513	0.61287
HNR	2.72E-02	9.27E-02	0.294	0.77176
alpha	-2.89E-01	7.61E-01	-0.379	0.70797
jitter	1.39E+03	6.16E+03	0.225	0.82407
Shimmer	4.12E-01	8.16E-01	0.505	0.6183

Experiment 1: Female Voices PC2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.03E+00	2.88E+00	1.4	0.17474
F0	9.16E-03	6.53E-03	1.404	0.17374
intonation	-3.25E-03	1.92E-03	-1.692	0.10412
dispersion	-7.57E-03	2.36E-03	-3.203	0.00395
glide	6.20E-04	1.59E-03	0.389	0.70081
HNR	-4.71E-02	8.66E-02	-0.543	0.59205
alpha	1.98E-01	5.92E-01	0.334	0.74121
jitter	-1.03E+04	7.46E+03	-1.384	0.17953
Shimmer	2.15E+00	9.74E-01	2.21	0.03733

Experiment 2: Male Voices PC1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.91E+00	3.35E+00	0.87	0.3931
F0	1.23E-02	6.43E-03	1.916	0.0678
intonation	-2.06E-03	3.28E-03	-0.628	0.5361
dispersion	-2.85E-03	2.37E-03	-1.204	0.2407
glide	6.48E-04	3.48E-03	0.186	0.8542
HNR	-1.85E-01	1.30E-01	-1.423	0.1683
alpha	-6.31E-01	4.34E-01	-1.452	0.1599
jitter	6.53E+03	3.70E+03	1.764	0.0911
Shimmer	6.57E-01	1.05E+00	0.625	0.5381

Experiment 2: Male Voices PC2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.81E+00	1.86E+00	-2.58	0.0167
F0	2.27E-02	3.58E-03	6.335	1.83E-06
intonation	-4.18E-03	1.83E-03	-2.287	0.0317
dispersion	-7.44E-05	1.32E-03	-0.056	0.9555
glide	-1.62E-03	1.94E-03	-0.836	0.4115
HNR	1.71E-02	7.25E-02	0.236	0.8154
alpha	6.72E-01	2.42E-01	2.781	0.0106
jitter	4.75E+03	2.06E+03	2.305	0.0305
Shimmer	-2.92E-01	5.84E-01	-0.499	0.6226

Experiment 2: Female Voices PC1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.02E+00	4.39E+00	1.371	0.18365
F0	-1.92E-02	9.97E-03	-1.926	0.06652
intonation	8.25E-03	2.93E-03	2.815	0.00983
dispersion	3.35E-03	3.61E-03	0.929	0.36233
glide	-6.59E-03	2.43E-03	-2.709	0.01253
HNR	-9.53E-02	1.32E-01	-0.721	0.47847
alpha	-1.86E+00	9.04E-01	-2.058	0.05109
jitter	-7.86E+03	1.14E+04	-0.691	0.49675
Shimmer	3.60E-01	1.49E+00	0.242	0.81101

Experiment 2: Female Voices PC2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.11E+00	2.09E+00	1.009	0.3235
F0	4.58E-03	4.74E-03	0.966	0.3439
intonation	-2.42E-03	1.39E-03	-1.735	0.0962
dispersion	-2.78E-03	1.72E-03	-1.622	0.1185
glide	2.98E-03	1.16E-03	2.573	0.017
HNR	-5.52E-02	6.29E-02	-0.878	0.3891
alpha	-2.25E-01	4.30E-01	-0.525	0.6048
jitter	-4.62E+03	5.41E+03	-0.854	0.4021
Shimmer	1.14E+00	7.07E-01	1.615	0.1199

Appendix F: Question 1 and 2 PCA

Question 1: Male Voices PCA

To examine Question 1, to determine the relations of ratings and acoustic measures across cultures and languages, the PCA here collapses the data gathered from ratings of male voices in one's native language, "Hello" in Scotland and "Hola" in Barcelona. A two dimensional solution was found, in which 85.55% of the variance was explained; PC1 explained 59.27% of the variance and PC2 explained 26.28%. The two personality traits that most strongly drove PC1 were trustworthiness (*r=0.90*) and confidence (*r=0.87*), with likeability having a very similar Pearson's r of 0.86. Dominance (*r=0.81*) and aggressiveness (*r=0.79*) were the main traits driving PC2. In *Figure 5* below, you can see the two dimensional solution for PCA for male voices for the original study (McAleer, Todorov, & Belin, 2014) and Experiment 1 combined.

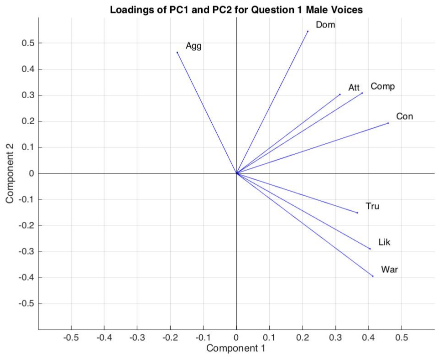

Figure 5

Question1: Female Voices PCA

Question 1 was also analyzed for females voices as well and a two dimensional solution was found, in which 81.91% of the variance was explained; PC1 explained 51.77% of the variance and PC2 explained 30.14%. The two personality traits that most strongly drove PC1 were trustworthiness (*r=0.93*) and likeability (*r=0.93*), with warmth having a very similar Pearson's r of 0.92. Dominance (*r=0.95*) and aggressiveness (*r=0.73*) were the main traits driving PC2. In *Figure 6* below, you can see the two dimensional solution for PCA for female voices for the original study (McAleer, Todorov, & Belin, 2014) and Experiment 1 combined.

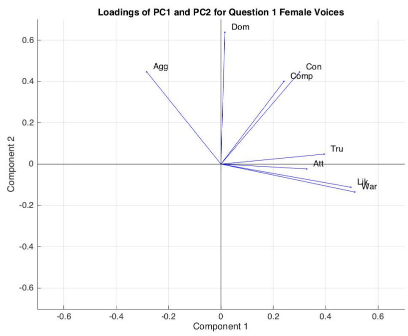

Figure 6

Question 2: Male Voices PCA

To examine Question 2, to determine the relations of ratings across cultures in the same language we performed a PCA, collapsing the data gathered from ratings of male voices in one's native language, "Hello" in Scotland and in one's foreign language "Hello" in Barcelona. Again, a two dimensional solution was found, in which 87.15% of the variance was explained; PC1 explained 57.61% of the variance and PC2 explained 29.54%. The two personality traits that most strongly drove PC1 were trustworthiness (*r=0.89*) and confidence (*r=0.85*) while dominance (*r=0.84*) and aggressiveness (*r=0.80*) were the main traits driving PC2. In *Figure 7* below, you can see the two dimensional solution for PCA for male voices for the original study (McAleer, Todorov, & Belin, 2014) and Experiment 2 combined.

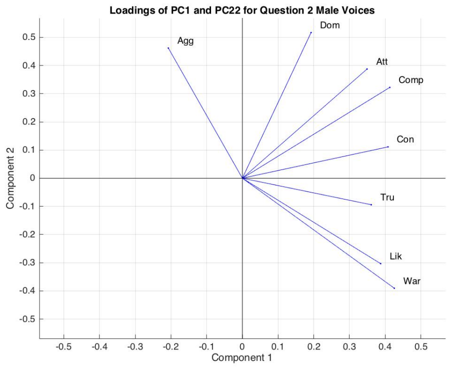

Figure 7

Question 2: Female Voices PCA

Question 2 was also analyzed for female voices in which a two dimensional solution was found, to explain 85.93% of the variance; PC1 explained 63.38% of the variance and PC2 explained 22.55%. The two personality traits that most strongly drove PC1 were trustworthiness (*r=0.94*) and likeability (*r=0.93*). Dominance (*r=0.83*) and aggressiveness (*r=0.80*) were the main traits driving PC2. In *Figure 8* below, you can see the two dimensional solution for PCA for female voices for the original study (McAleer, Todorov, & Belin, 2014) and Experiment 2 combined.

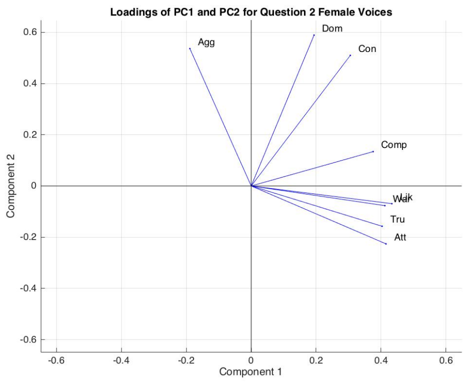

Figure 8