Vocal Typology in Commerce: Deployment and Employment

Tyler McPeek

Abstract

Individual voices are not uniformly similar to others, even when factoring out speaker characteristics such as sex, age, dialect, and so on. Some speakers share common features and can cohere into groups based on gross vocal similarity. A study was performed to generate a taxonomy based on these "voice types." Perceived similarity judgments of voice pairs using a database of 100 female and male American English voices were collected and submitted to a hierarchical clustering analysis to generate the initial groupings of individual voices into types, separately for female and male voices. These types, in turn, were labeled based on auditory judgments by expert listeners on nominal scales (e.g., voice quality, mean pitch, pitch variability, and speaking rate) as well as an initial acoustic analysis using automated measures. The new typology revealed a total of 9 female and 9 male voice types, with voice quality, mean pitch, and pitch variability playing the largest roles in determining the taxonomy for both sexes. This new vocal typology will find utility in academia (phonetics, discourse, sociolinguistics, genetics, and other fields), forensic linguistics, public and private sector business and marketing, voice acting, and general public interest arenas. In this paper, the study, resulting vocal typology, and commercial applications are summarized. Voice types and listeners' emotional perceptions of these types can be used to both deploy voice types as a means of more effectively selling products or services, as well as to apply voice types and associated, perceived personality qualities based on the types to best employ individuals for different types of tasks that bring them into contact with public consumer and company client listener perceptions.

Introduction

Both linguistic and non-linguistic information are transmitted in the same signal of a spoken voice, and the correct identification of linguistic information can be crucially dependent on the acoustic variability associated with non-linguistic, or indexical information. The indexical properties of speech specify information about the history and/or current state of the speaker him/herself, such as sex, age, dialect, emotion, fatigue, pathology and, most relevant to this study, speaker identity. Individual speakers can be identified by numerous means, both linguistic and nonlinguistic, but for these purposes, speaker identity refers to all aspects of the speech signal that are independent of other indexical and linguistic properties. These can be a product of the vocal and/or speech anatomy of the individual as well as idiosyncratic physiological patterns. Prior work in the correlates of Speaker Identification (SPID) by human listeners have identified such cues as speaking fundamental frequency (Atal, 1972; Iles, 1972; Jassem, et al., 1973; LaRiviere, 1975;

Mead, 1974), mid to high frequency spectral information, such as higher formants F3−F5 (Goldstein, 1976; Jassem, 1968; Iles, 1972; LaRiviere, 1975), nasality (Glenn & Kleiner, 1976; Su, et al., 1974), temporal speech features (Abberton & Fourcin, 1978; Johnson, et al., 1984), voice quality (Hollien & Majewski, 1977; Johnson, et al., 1997; Zalewski, et al., 1975), fricative articulation (Ingemann, 1968; Schuartz, 1986) -- and others (Hirson & Duckworth, 1995; Lass, et al., 1976; Orchard & Yarmey, 1995; Wolf, 1972; Young & Campbell, 1967). Anatomical and physiological features of the speaker are static to some extent, determined by factors such as the size and health of the vocal folds, air flow volume from the respiratory system, length, width, and shape of the vocal tract, the sufficiency of the velopharyngeal port, as insufficient closure of the velum may result in more nasality, and dentition. In addition, the way in which we use our articulators can also be unique, exhibiting more or less of various characteristics in speaker styles, such as degree of coarticulation and fast or slow speech (Hollien, 2002). Interspeaker variability in both speaker characteristics based on articulator usage and anatomically fixed features of speakers are likely to inform the groups in a typology of the type revealed by the present study.

Individual voices are not uniformly similar to others, even when factoring out speaker characteristics such as sex, age, dialect, and so on. Some speakers share common features and thus may naturally form groupings, termed hereafter as "voice types," that have not been systematically described or labeled prior to this study on "vocal typology" (McPeek, 2013). In both human perception of speaker identity and in machine-based approaches, error rates are usually above zero, permitting the examination of patterns of confusion *among* individual voices. In all cases, the error patterns do not show a random distribution of errors across competing voices. Rather, a given voice tends to be more confusable with one or more of a limited number of other voices in the test set. For example, in a classic study, Bricker and Pruzansky (1966) examined the effect of stimulus duration and type on the identification of familiar male voices. In the course of doing so, they generated confusion matrices for the ten voices under each experimental condition. In all cases, when voices were misidentified, they were not confused at equal likelihood with all other nine voices in the test set. Instead, each voice was typically confused with 1 or 2 others consistently (although, asymmetries in the matrices were also observed). A reanalysis of the error rates in the monosyllabic condition of the study shows that the most confusable voice for each of the ten talkers garnered 27%−53% of the errors recorded, well above a chance level of 11%. In other words, voices in the set were not uniformly similar: some were approximate sound-alikes or showed high degrees of similarity. Such nonrandom error patterns in this and other studies serve as evidence that voices may naturally cohere into vocal similarity groups in the ear and mind of listeners.

That voice types naturally occur is not especially surprising. Speaker identities can be confused over a phone or in other degraded listening conditions. Folk terms exist for vocal qualities that are not necessarily pathological but are distinctive, such as "nasally," "whiny," "gravelly," "droning," "staccato," and others. What remains, however, is a systematic approach for identifying the number and type of the most common vocal stereotypes, or types, that speakers cohere into based on human perception. An inventory of voice types can be developed which is independent of other speaker characteristics (e.g., age, sex, dialect, pathology) and which serves to reduce the vast population of speaker identities by voice into a more manageable taxonomy of common types. Such voice types may play a role, as do other indexical properties, as perceptual units that partly influence the processing of linguistic and nonlinguistic information by human listeners. Their existence also points to numerous applications. In the forensic domain, SPID is a very common analysis required of audio evidence in civil and criminal cases and, yet, the duration of the speech samples and their quality can often preclude a highly confident judgment of the match/mismatch to the voice of a defendant or a relevant party in a case. However, such evidence recordings may be of sufficient caliber to permit a match/mismatch determination on the basis of a more gross category, such as a voice type. The evaluation of voice talent is also a growing field, given the increasing use of digital animation in the enter-

tainment industry, spoken word in digital broadcast, and other factors. While individual vocal attributes such as "pleasantness" or "authority" have been examined in prior work (Beebe-Center, 1965; Oyer & Trudeau, 1984; Bugental & Lin, 1997; and others), there is currently no rubric or automated procedure for classifying all of the relevant characteristics of a talented voice. Voice talent could be fit into a voice type taxonomy for increased ease of identifying the proper vocal talent for a given commercial application. This would include public service announcements and advertisement narration, where vocal pleasantness correlates such as trustworthiness, sex appeal, and overall pleasantness or friendliness play a significant role in listener impression, attention to message, and overall decision making and effectiveness of the message. Employment screening based on vocal typology is also an application that is already being utilized for commercial purposes based on individual indexical property algorithms, but could be vastly improved through screening based on an existing typology. The positing of a voice type taxonomy ultimately serves to reduce the vast number of speaker identities within a given sex/age/dialect subpopulation down to a manageable and useful number of categories and indexical profiles. Prior to this study, no taxonomies of perceived voice types had been proposed for male and/or female voices in any language.

Resulting Voice Type Taxonomy

In this study, voice types were developed through extensive similarity judgments of pairs of voices from a large database that represents healthy young and middle aged voices speaking American English. These judgments were submitted to different types of statistical analyses, including a Hierarchical Clustering Scheme (HCS) analysis, to yield a taxonomy of voice types. For female and male voices, nine types of voices each were discovered. Each of these 18 types has its own unique acoustic signature, with characteristics that were selected on the basis of prior studies of the acoustic correlates of individual speaker identification, then submitted to an auditory analysis by two phonetically-trained expert listeners and a machine-assisted acoustic analysis focusing on pitch, voice quality, prosody, and articulatory cues to speaker identity. The results are represented in Tables 1 and 2 in terms of a simple coding system for data reduction purposes:

- Speaking Rate: S(low) vs. F(ast) — Relative to the average rate observed in the database
- Voice Quality: R(ough) vs. C(lear)
- Mean Pitch: H(igh) vs. L(ow) — Relative to the average level observed in the database
- Pitch Variability: M(onotone) vs. D(ynamic) Relative to the average range observed in the database
- Articulatory Effort: E(nunciated) — Degree of hyperarticulation relative to the range observed in vowel space area
- Nasality/Orality: N(asal) — Degree of atypical nasality in speech relative to normal utterances

From the judgment of expert listeners, five speech and voice characteristics were required to uniquely identify these types, and the same set proved viable for both genders: voice quality, articulatory effort, mean pitch, pitch variability, and speaking rate. The two taxonomies differed from one another in terms of the relative importance of these characteristics, the heavy reliance of female voices on fewer characteristics, and most importantly (Harnsberger and McPeek, 2013), their relative independence of a related indexical property of speech, vocal age. Male voice types used all five characteristics liberally, although a slow speaking rate characterized the two largest types. Most male voice types also consisted of a mix of voices of different vocal age, defined chronologically or perceptually (perceived age data available for this database from Harnsberger, et al., 2010), while vocal age appeared highly salient among female voices, which were divided into multiple types representing greater shares of the database. For female voices,

Table 1. Female Voice Types

Voice Type	Size	SR	VQ	MP	PV	AE	N/O
F-V1	28%	F					
F-V2	6%		R	L			
F-V3	4%					E	
F-V4	12%						
F-V5	6%	F	R	H			
F-V6	34%			H			
F-V7	4%	S			D		
F-V8	4%	S		H		E	
F-V9	2%	F	C	H		E	

Table 2. Male Voice Types

Voice Type	Size	SR	VQ	MP	PV	AE	N/O
M-V1	34%	S	R		D		
M-V2	8%		R	L			
M-V3	4%	F	R	L	M	E	
M-V4	6%			H	M		
M-V5	24%	S		H			
M-V6	8%	F	R	L	D		
M-V7	10%		R	H			
M-V8	2%	F				E	
M-V9	4%	S			M		

chronological age influenced distribution of voices into types, with younger voices grouped separately from middle aged voices, while male voice types were more heterogeneous with respect to age. Vocal age appears more salient in judging female voice type, and this observation is congruent with higher accuracy rates in estimating female vocal age in prior studies (Schotz, 2006; Hughes & Rhodes, 2010). Vocal age perception was not a focus of this study, but was explored to ascertain that vocal similarity judgments by listeners were not based purely on speaker age.

The present research and experimentation has yielded a working set of voice types for American English, based on the above described type of similarity judgments by untrained listeners. This research has tapped directly into the human's intuitive understanding of voices inherent non-uniform similarity to one another. Data reduction and acoustic analysis were then used to cull, sculpt, confirm, and label these natural class voice types into a working typology of American voice types separately for females and males. Use of a larger database or larger sample of the listener population might result in nominal changes to this typology, without changing the fundamental aims and results of the project—which were to build a natural class themed taxonomy of vocal stereotypes that is based on the intuitive judgments of the untrained hu-

man ear. The untrained listener remains as the most reliable mechanism for human voice identification and this fact is a foundation for this study.

Classifying Voice Types by Emotional Perception

So, are you a Barry White or a Gilbert Gottfried, a Marilyn Monroe or a Shirley Temple, a droner or a diva, an authoritarian or a nurturer? How similar or dissimilar is your voice to other voices and how might that affect how you are perceived by those around you? Are you uniquely suited by the anatomically motivated characteristics and acoustic cues associated with your voice type to be a policeman, a public safety announcer, a voice talent actor, a preacher, a teacher, or some other occupation? These are the types of questions that flow from the study of vocal typology.

A list of vocal descriptors has been compiled which can be utilized by expert analysts in the final labeling of the voice types yielded from the matrix. These descriptors have been selected from a compilation that draws primarily from five over-lapping groups: descriptors used by speech pathologists and speech experts, linguistic descriptors, descriptors used to type and specify voices by singing voice experts and voice talent specialists, common-use descriptors used by non-experts to describe the qualities of a voice, and common-use descriptors used by non-experts to describe the emotional response of listeners to a given voice. A representative sample of descriptors and descriptive vectors/scales from each group are shown in

Table 3. Vocal Descriptors by Category

Speech Pathology	Linguistic	Singing Voice	Popular/Common Use General Characteristics/Voice Quality	Emotional Response
rough	creaky	timbre	gravely	whiney
breathy	laryngeal	register	shrill	droning
strain	breathy	baritone	scratchy	commanding
hoarse	murmured	chocolate	deep	nerdy
clear		velvet	high	meek
flexibility		soaring	low	strong
asthenia		range	husky	weak
nasal resonance		dark	monotone	authoritative
loudness		bright	child-like	masculine
pitch range		ringing	musical	feminine
pitch		nasal	clear	flighty
quality		rich	rough	annoying
rate		rough	raspy	bright
prosody		light	throaty	upbeat
glottal fry		heavy	smooth	creepy
diplophonia	bass	gruff	unsure	
		tenor	grating	golden
		alto	halting	sexy
		soprano	full	reassuring

Table 3.

Descriptors used to describe emotional responses to certain voice types are useful; however, they constitute a different method of labeling a voice type that is most efficiently yielded from further testing on the established voice types by un-trained, non-expert listeners, utilizing an emotional response rating system similar to those used in previous vocal pleasantness experimentation.

Five hypothetical voice types (sex specific, one set for each sex) might look like those outlined in Tables 4, 5, and 6.

The labels in Table 4 would be considered the "descriptor labels." These labels could be assigned corresponding labels by arbitrary coding, avoiding emotional, social, or pathology based assumptions about a voice type or speaker by label (until such associations can be made and labeled to each type by further vocal pleasantness studies), as shown in Table 5.

Celebrity monikers judged by experts to be included in and able to represent a given type could then be assigned to make the labels more interesting, relevant, and easier to remember for the general public, as shown in Table 6. This type of labeling system also risks emotional and social associations with each type that might be better avoided for scientific labeling, but nonetheless carry commercial and public interest appeal for general use.

General public labeling can also utilize emotional response labels in a personality type labeling system of the type shown in Table 7, if appropriate for the given application of the typology. These types of labels

Table 4. Perception Based Descriptor Labels

Scratchy Nasally
Deep Baritone
High Melodic
Low Breathy
Rapid Staccato
Dark Ringing
Bright Heavy
Rich Deep
Low Monotone
Rough Halting

Table 5. Arbitrary Coding Type Labels

Type 2A
Type 4C
Type 1D
Type 7S
Type 3R
Type 6T
Type 7Y
Type 8D
Type 5B
Type 0Y

Table 6. Celebrity Moniker Labels

"Gilbert Gottfried Type"
"George Clooney Type"
"Shirley Temple Type"
"Marilyn Monroe Type"
"Lou Costello Type"
"Casey Kasem Type"
"Barry White Type"
"Julia Childs Type"
"James Earl Jones Type"
"Woody Allen Type"

Table 7. Emotional Response Based Personality Type Labels

"*The Whiner*"
"*The Golden Voice*"
"*The Kid*"
"*The Preacher*"
"*The Seducer*"
"*The Droner*"
"*The Mouse*"
"*The General*"
"*The Thinker*"
"*The Reluctant Participant*"

also can be enhanced by future research done on this typology and vocal typology generally by researchers in the field of vocal pleasantness and sociolinguistics. Again, the application of these types of labels generally carries tradeoffs via public interest vs. scientific application. In both cases, they may be better applied when they can be justified properly by listener experimentation focused on emotional and favorable vs. unfavorable response to each individual voice type on a given set of scales or of one voice type as compared to another, on a given set of dichotomous criteria, such as general pleasantness, trustworthiness, authority, sex appeal, perceived intelligence and intellectual prowess, and so on.

Personality type labels in particular give dimension and emotional association to the types that can provide easy mental association and quick allusion to the general characteristics and sound of the voices in a given type, as well as hinting at the voice type's commercial viability for specific purposes.

Commercial Applications

Application and Value of Research

While this study was originally performed for purposes of reporting experimentation into speaking voice types and an analysis of those results with their implications for our current understanding of natural vocal typology, it is desirable that this new typology of American speaking voices might find opportunity for further definition through future study. These studies might explore cross-linguistic and general universality of voice types, vocal pleasantness correlates, and others. Most especially, however, practical application outside of academia is desirable to the researchers. Currently, there are companies operating in the private sector that claim to employ algorithms that can interpret how a voice makes others feel, for purposes of choosing an employment candidate's suitability for particular types of work (Shahani, 2015). These companies screen candidates by voice, on behalf of their clients, to determine whether a candidate is the best fit for the duties required of a given employment position (usually telephone customer service). These companies' algorithms are opaque and controversial, and moreover, do no employ vocal typology in their criteria. Voice types and vocal typology as a concept should find potential revision through future experimentation and eventual utility in the public, academic, and private sectors. What follows is a partial listing of possible applications for vocal typology across an array of fields and sectors.

Academic and Scientific Study

Other phoneticians, acoustic experts, and linguists might seek to perfect or recast these types, leading to a more universal typing system, to be of academic and scientific utility to a broad range of different fields. Geneticists might seek to explore whether genetically related people share common voice types separate of environmental factors, such as dialect. Government bodies might seek to use this system to determine which voice types are most pleasant or well-received for different types of public service announcements, warning systems, or automated, verbally-administered testing procedures. Cross-linguistic as well as unified voice types might be used to determine regional typology trends in acoustic signals of voices, to complement existing linguistic typology research on the vast array of world languages.

Forensic

The science of SPID and its different methods have long been plagued by issues of degree of accuracy. Both in and out of the courtroom, forensic acousticians and phoneticians have had their work fall under criticism for the degree to which they are able, or not, to make a positive match or to rule out a voice through investigation and analysis. One of the issues with acoustic as well as perceptual-aural analyses have been their estimated degree of accuracy by percentage. While the usefulness of traditional SPID is apparent, it's admissibility in a court of law, where lives often hang in the balance, has been under scrutiny

for years. In light of the high degree of accuracy in identification methods such as DNA matching in recent years, it is not surprising that judges and the judicial system might look unfavorably on scientific identification methods that yield a shaky 60% match or often even less. Adding to the skepticism of the judicial system, is the "CSI Effect" exhibited by juries in recent years.

Though blood typing in the absence of DNA evidence may only produce evidence that the defendant and the perpetrator of a crime share a blood type belonging to hundreds of millions of others in the world, it is admitted freely into the courts—as in the case of blood typing, a match is a near certain match, regardless of the size of the member population in the group to which the match belongs. So, it would seem that for the courts, and often the public at large, a 100% match to a type belonging to the defendant and millions of others would be preferable to a less certain, though far more personalized match, as in the case of an SPID match estimation made by a forensic linguistics professional. Additionally, despite ample experimental and academic evidence suggesting the fallibility of voice line-up and other SPID matches by witnesses, both the judicial system and juries have traditionally looked favorably upon such evidentiary presentations in the courts. What seems to be missing is an alternative identification method that might pass muster with the contemporary expectations of judges, juries, and professionals—one that would be neither too stringent to raise accuracy issues, nor overly broad enough to be lost in futility. If a reliable vocal typing system were available to the courts and forensic linguists, it would be of great utility.

Government and Public Interest

Of course, voice typing would yield not only a possible greater application in the forensic/judicial realm, but would also have the advantage of opening up the field to a new and wide range of civilian, corporate, and government uses—where its application might be less controversial. The narrow criminal application outlined above is only one of thousands of possible applications, including everything from marketing and advertising to warning systems and public service, safety, and welfare announcements. In addition, vocal types, which naturally include celebrity voices and are accessible to the public, would have commercial viability and could bring much needed resources and attention to the field of forensic phonetics and linguistics generally. Perception of ones voice by others, and the implications involved therein, would be valuable information to individuals, corporations, and political entities alike. Online and elsewhere, services that offer the visitor the opportunity of having their voice typed would be of great public interest. Individuals could learn what the characteristics of their voice type are, as well as what famous voices they resemble typologically and how they might be perceived by others at work and in social settings, based on their voice. In addition, individuals could experiment to see if their voice belongs to the same type as their loved ones and others, opening the door to interesting speculation on how couples and friends might be drawn together in part by voice, and whether or not voice is passed down genetically and/or by proximity and nurturing of ones children.

Business and Marketing

Business and Marketing applications abound for a viable voice type system. Determining which voice type would be most effective for marketing of specific products and services (trustworthy for insurance and financial products, exciting or relaxing for travel services, etc.) through marketing research, would in the end be of shared interest to academics looking to determine the personality qualities associated with certain indexical properties of voice co-assigned to each type, as well as commercial marketers. This would also be useful to voice casting agents to provide their clients with short lists of available voice actors filtered by requested type or desired quality. Voice type could become an indispensable element of an actor's resume. How this research might be of value to singing voices and jingles is also an open and intriguing question that would require further research. Marketing research entities interested in creating personality-reflective

voice types for narrative marketing services to specific industries could provide valuable private sector funding and research partners for vocal typologists interested in vocal pleasantness and voice perception in sociologically-geared experiments on telephone discrimination, linguistic profiling, and other areas of inquiry.

Sociolinguistics and Discourse

How ones voice is perceived in discourse and daily human interaction has many implications for how a person is treated in society. The study of how certain voice types are perceived via their indexical properties in the speech signal and how this perception affects their overall treatment in discourse would be of interest to sociolinguists and discourse analysts. Issues of attraction, repulsion and other emotional responses based on the physical properties of the speech signal would be an important overlap to the study of vocal pleasantness and discourse analysis within and between vocal types.

Conclusions

As discussed above, there are ample opportunities to increase both the understanding and utility of vocal typology through further research and experimentation. In the current study, one set of identical twins was observed superficially by two expert listeners for voice similarity. This observation was not acoustically measured, but consisted instead of a short interview only. Generally, the twins were observed to have a remarkably similar voice type, with their voices often being mistaken over the telephone and in other environments. These interviews were not scientific enough to yield publication data, but it did open the door to a wider study of people who share similar body types or genetically identical vocal tract constructions to be studied for vocal similarity.

It is also the intention of the current author to complete a follow-up study on native speakers of Japanese. A study of this type would help determine 1) to what degree voice types are universal and cross-linguistic in nature and 2) how valid the results of this study are when replicated with a different population of speakers and/or listeners generally.

These types of further studies seek to increase the academic understanding of vocal typology through continued research. Equally exciting would be opportunities to collaborate with commercial entities like market researchers and human resource specialists to test a typing system for incoming voices. Commercial viability for voice typing will ultimately determine the scope of research funding available towards improving and expanding this new scientific field.

References

Abberton, E. & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and Speech, 21*, 305−315.

Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *The Journal of the Acoustical Society of America, 52*, 1678−7697.

Beebe-Center, J. Gilbert. (1965). *The psychology of pleasantness and unpleasantness.* New York: Russell & Russell.

Bricker, P. D. & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America, 40(6)*, 1441−1449.

Bugental, B. B. & Lin E. K. (1997). Attention-grabbing vocal signals: Impact on information processing and expectations. *Personality and Social Psychology Bulletin*, 23(9), 965−973.

Glenn, J. W. & Kleiner, N. (1976). Speaker identification based on nasal phonation. *The Journal of the*

Acoustical Society of America, 43, 368−372.
Goldstein, U. G. (1976). Speaker-identifying features based on formant tracks. *The Journal of the Acoustical Society of America 59*, 176−182.
Harnsberger J. & McPeek, T. (2013). Towards a typology of American English. Proceedings from *Meeting on Acoustics: ICA 2013*. Published by the Acoustical Society of America through the American Institute of Physics, 19, 060156.
Harnsberger, J., Shrivastav, R., Brown, W. (2010). Women's vocal aging: A longitudinal approach. *Proceedings of Interspeech*. 466−469.
Hirson, A. & Duckworth, M. (1995) Forensic implication of vocal creak as voice disguise. BEIPHOL, *Studies in Forensic Phonetics, 64*, 67−76.
Hollien. H. (2002). *Forensic voice identification*. London, England: Academic Press.
Hollien, H. & Majewski, W. (1977). Speaker identification by long-term spectra under normal and distorted speech conditions. *The Journal of the Acoustical Society of America, 62*, 975−980.
Hughes, S. & Rhodes, B. C. (2010). Making age assessments based on voice: The impact of the reproductive viability of the speaker. *Journal of Social, Evolutionary, and Cultural Psychology, 4*, 290−304.
Iles, M. (1972). Speaker identification as a function of fundamental frequency and resonant frequencies. (Unpublished doctoral dissertation). University of Florida, Gainesville.
Ingemann, F. (1968). Identification of the speaker's sex from voiceless fricatives. *The Journal of the Acoustical Society of America, 44*, 1142−1144.
Jassem, W. (1968). Format frequencies as cues to speaker discrimination. *Speech Analysis and Synthesis, 1*, 9−41.
Jassem, W., Steffen-Batog, M., & Czajka, S. (1973). Statistical characteristics of short-term average Fo distribution as personal voice features. *Speech Analysis and Synthesis, 3*, 209−228.
Johnson, C. C., Hollien, H., & Hicks, J. W. (1984). Speaker identification utilizing selected temporal speech features. *Journal of Phonetics, 12*, 319−326.
Johnson, K. & Mullennix, J. W. (1997). *Talker variability in speech processing*. Waltham, MA: Academic.
LaRiviere. C. L. (1975). Contributions of fundamental frequency and formant frequencies to speaker identification. *Phonetica, 31*, 185−197.
Lass, N., Hughes, K., Bower, M., Waters, L., & Bourne, V. (1976). Speaker sex identification from voiced, whispered and filtered isolated vowels. *The Journal of the Acoustical Society of America, 59*, 675−678.
McPeek, T. (2013). American voice types: Towards a vocal typology for American English (Doctoral dissertation). Retrieved from ProQuest. (3586411)
Mead, K. (1974). Identification of speakers from fundamental frequency contours in conversational speech. Joint Research Unit Report 1002, 1−22.
Orchard, T. L. & Yarmey A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology 9(3)*, 249−260.
Oyer, H. J. & Trudeau, M. D. (1984). Temporal speech characteristics and vocal pleasantness in the aging voice. *The Journal of the Acoustical Society of America, 75(S1)*, S41−S41.
Schuartz, M. F. (1986). Identification of speaker sex from isolated voice fricatives. *The Journal of the Acoustical Society of America, 43*, 1178−1179.
Schötz, S. (2006). Perception, Analysis and Synthesis of Speaker Age (Doctoral dissertation). Retrieved from Lund University Publications. (25959)
Shahani, A. (2015). How algorithms are deciding whom to hire, based on voice. All Tech Considered. Retrieved from http://www.npr.org/sections/alltechconsidered/2015/03/23/394827451

Su, L., Li, K., & Fu, K. (1974). Identification of speakers by use of nasal coarticulation. *The Journal of the Acoustical Society of America, 56*, 1876−1882.

Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America, 51*, 2044−2055.

Young, M. A., & Campbell, R. A. (1967). Effects of context on talker identification, *The Journal of the Acoustical Society of America, 42(6)*, 1250−1254.

Zalewski, J., Majewski, W., & Hollien, H. (1975). Cross-correlation of long-term speech spectra as a speaker identification technique. *Acustica, 34*, 20−24.