



Original Research

American telephone voice: Socio-phonetic features of identity in professional and non-professional discourse

by Tatiana I. Shevchenko and Tatiana V. Sokoreva

Tatiana I. Shevchenko Moscow State Linguistic University, Russia tatashevchenko@mail.ru

Tatiana V. Sokoreva Moscow State Linguistic University, Russia jev-t@yandex.ru

Article history Received November 14, 2022 | Revised May 10, 2023 | Accepted June 1, 2023

Conflicts of interest The authors declared no conflicts of interest

Research funding No funding was reported for this research

doi [10.22363/2521-442X-2023-7-2-72-79](https://doi.org/10.22363/2521-442X-2023-7-2-72-79)

For citation Shevchenko, T. I., & Sokoreva, T. V. (2023). American telephone voice: Socio-phonetic features of identity in professional and non-professional discourse. *Training, Language and Culture*, 7(2), 72-79.

The study is concerned with the role of socio-phonetic characteristics of American speakers' voices in identifying gender and age identity in professional and non-professional communication. We have to develop and test the methodology previously applied in such specific areas as corpus analysis, forensic phonetics, multi-modal research, and intraspeaker variability to address the problem of identifying personality by voice. The aim of the research is to view variations in the prosodic forms of American English speakers and examine age-related and gender-specific prosodic features in telephone communication. The study intends to address the following research questions: How do the following prosodic features contribute to distinguishing age- and gender-related changes in the human voice: mean pitch, maximum pitch, minimum pitch, pitch range, mean intensity, jitter, shimmer and harmonics-to-noise ratio? Which combinations of features are associated with young, middle-aged, and senior voices of men and how much are they different from or similar to women's voices? The auditory and acoustic computer analyses were conducted on American English dialogues with 30 speakers, equally balanced for gender (5 men and 5 women in each age group) and three age groups (young, middle-aged, senior), taken from the American telephone speech corpus to measure the prosodic parameters of pitch, pitch range, intensity, jitter, shimmer, and harmonics-to-noise ratio supported by statistical data processing in Minitab programme. The apparent-time technique of data presentation and the comparative analysis allow discovery of the dynamics of voice changes over time which could be later applied to personality identification. Comparing the data based on eight parameters' values in three age groups balanced for gender we have found that although all the selected measurements proved to be relevant for either gender group, the age-related trajectory of voice dynamics may be different in male and female voices regarding their specific prosodic characteristics. Prosodic features of mean pitch, maximum pitch and minimum pitch have been confirmed to define both individuals and groups of people of a certain age and gender and could be considered as both speaker-identifying and group-identifying characteristics. Voice quality features, apart from being previously found to signify the emotional states of speakers, may also be observed to characterise certain age and gender groups; hence, they may also serve as speaker-identifying characteristics.

KEYWORDS: *telephone voice, prosody, age, gender, American English, socio-phonetic features*



This is an open access article distributed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which allows its unrestricted use for non-commercial purposes, subject to attribution. The material can be shared/adapted for non-commercial purposes if you give appropriate credit, provide a link to the license, and indicate if changes were made.

1. INTRODUCTION

In the present study, the human voice is applied as a cover term for a complex of prosodic characteristics which indicate the social categories of age, gender, and style. Thus, the sociolinguistic perspective provides for a fair account of the voice impact which is due to concrete features of pitch, pitch range, intensity and voice quality measured objectively by a computer programme. These features are utilised by the community in a number of professional and non-professional types of discourse. The areas which could be specified are:

- public speaking (politicians, for instance, tend to have voice coaches or/and image makers who instruct public figures how to use their voice effectively to influence the audience);
- acting (professional training of actors is a vital means of projecting an image in building up the desired characters. The chances are that voicing in films, especially dubbing, is made by different actors for the purpose);
- news reading in mass media (the particular presentation of the message, especially the commentary, has to sound impartial and objective for which a certain voice quality is necessary);

- singing (singing requires a long-term practice and care of the performer's voice in special musical institutions);
- business (training staff for call centres and other ways of promotion in commercials needs voices with social attractiveness and the gift of persuasion; the voices are tested and rated for intelligibility and social acceptability through surveys);
- forensic phonetics (in criminology the recorded voice of an individual may serve as evidence on the grounds of forensic speech analysis supported by objective facts and the expert's decision);
- health service (voice pathology caused by age, injury and disease is treated in medicine, as the environmental changes of the present century keep aggravating the situation despite the introduction of new technologies for the sake of human communication);
- academic activity, lecturing in particular, has been traditionally associated with the voice of a learned professional whose impact is partially based on the voice characteristics;
- Internet-mediated communication in social network may include video and audio messages which need to be identified as belonging to certain people; Internet-mediated communication through various gadgets appears to be displacing traditional telephone communication;
- face-to-face interaction of people presents the most common means of oral speech discourse where the human voice is only part of the multimodal communication.

The present study is focused on voice features in terms of prosody which were found in telephone recordings. We claim that these features will be equally well-employed both in professional and non-professional communication. The complex of prosodic features will facilitate the recognition and evaluation of particular individuals. To pursue this aim, the study will address the following research questions: How do prosodic features such as mean pitch, maximum pitch, minimum pitch, pitch range, mean intensity, jitter, shimmer, and harmonics-to-noise ratio contribute to distinguishing age- and gender-related changes in the human voice? Which combinations of features are associated with young, middle-aged, and senior voices of men as different from or similar to women's voices?

2. MATERIAL AND METHODS

Sound material for analysis was taken from free online resources of Linguistic Data Consortium, Switchboard-2 Telephone Speech Corpus, a collection of about 2400 two-sided telephone conversations among speakers from all areas of the United States (Godfrey et al., 2021). The Switchboard Corpus represents the telephone voice database where the speakers agreed to a selected topic and to being recorded. Nobody spoke twice with the same interlocutor and on the same topic. Thus, the speaking conditions under control provide for emotionally neutral, calm, friendly type of speech. There is no comment on the speakers' voice pathologies in the corpus annotation. The technical recording characteristics: sound sample type is 2-channel ulaw with an 8000 Hz sample rate.

For the present research the samples of 30 American English speakers, equally balanced for gender (5 men and 5 women in each age group) in three age groups (10 young adult speakers aged 20-39; 10 middle-aged adults aged 40-59; 10 senior speakers aged 60-69) were selected. By presenting three generations of speakers (young, middle-aged, senior) we observe the human voice characteristics in the *apparent time*, i.e., in the dynamics of life-time development.

The style of telephone conversation is considered to be representing both professional and non-professional registers. We focused on the basic features of pitch, pitch range, intensity and voice quality.

The methods used in the study include the acoustic analysis of the speech signal using the PRAAT computer programme (Boersma & Weenink, 2017) aimed at identifying the following measurements of the data: mean pitch – F0 mean (Hz); minimum pitch – F0 min (Hz); maximum pitch – F0 max (Hz); pitch range – F0 range (st); mean intensity – Int mean (dB); jitter (microsec); shimmer (%); harmonics-to-noise ratio – HNR (dB).

All the measurements were taken from the accented syllables in the phrases. While calculating pitch and intensity parameters, some of PRAAT's values were hand-corrected to remove pitch tracker errors. Low F0 values, typically a result of a creaky voice, were excluded from the analysis. The pitch range values were converted from Hertz into semitones by means of the Semitone online converter.

Jitter is a measure of the irregularities in the frequency at which the vocal folds vibrate. It shows the deviation from true periodicity of a presumably periodic signal of a sound wave. The present study appeals to the parameter of jitter (local, absolute) that is the average absolute difference between consecutive periods calculated in seconds, namely in microseconds – μs (1000 microseconds equal to 1 millisecond). The threshold for pathology is claimed to be 83.200 μs (0.083 ms) according to certain studies (Deliyski, 1993).

Shimmer, a cycle-to-cycle variability of the period amplitude of vocal folds vibration, is used as one of the measures showing the micro-instability of the voice. The present research resorts to shimmer (local) that is the average absolute difference between the amplitude of consecutive periods, divided by the average amplitude, measured in percentage points. The threshold for pathology is stated to be 3.81% (Deliyski, 1993).

Harmonics-to-noise ratio is an acoustic parameter used to estimate the level of noise in human voice signal, measured in decibels (dB). The HNR of 0 dB means that there is equal energy in the harmonics and in the noise (Boersma & Weenink, 2017).

The statistical analysis of the obtained data was realised by finding the median values of all the parameters as well as conducting a Kruskal-Wallis statistical test in order to trace the significant differences in the above-mentioned parameters, the factors being age (three levels: young, middle-aged, senior) and gender (two levels: male and female). The received data were also subjected to a comparative analysis aimed at revealing age-related and gender-specific characteristics in the analysed speech

samples. The statistical analysis together with graphic representation of the results were performed using Minitab and Excel programmes.

3. THEORETICAL BACKGROUND

3.1. Previous research

When two people start a telephone conversation their first words may indicate a lot of socially relevant information about their age, gender, social status and where they come from. Thus, the interlocutor's voice may either help to identify an old acquaintance or give the first impression about a new one, thanks to its segmental and prosodic characteristics (Shevchenko & Sokoreva, 2019). Listeners judge the age and gender of speakers on the basis of both prosodic and segmental cues fairly accurately (Helfrich, 1979).

Prosodic characteristics are reported to have a constant decline with ageing, which concerns fundamental frequency and intensity levels, articulation slowing down and voice quality changes (Linville, 2001). Along with a physiologically determined drop in vocal tract capabilities the prosody of an ageing voice is prone to constant development caused by the necessity to maintain word accentual prominence, intelligibility, and clarity of speech.

The aged-related changes embrace all the prosodic characteristics of human speech: duration, fundamental frequency, and intensity, with gender-specific modifications of the general trend. In view of the constant loss of high-pitched registers in elderly people's speech the accentual prominence, essential for speech intelligibility, is achieved by means of greater variance in the prosodic characteristics of accented syllables. The parameters which proved significant in that respect in telephone voices across the three generations are: increase in the accented syllables duration with age which is a sign of slower tempo and compression of unaccented syllables; increase in maximum intensity of accented syllables compared to unaccented syllables intensity values; increase in pairwise variability indices for maximum fundamental frequency values which is indicative of greater variance of the high pitch values in older people's speech (Shevchenko & Sokoreva, 2018).

Thus, prosodic prominence in an ageing voice is affected by pitch, intensity and duration that contribute to keeping up the word accentual patterns adding to intelligibility and clarity of speech despite the naturally occurring changes in vocal anatomy and physiology.

Among the personality features in prosody, rhythm is not the least in providing for speech comprehensibility and clear enunciation. The stress-timed rhythm of English based on accentuation turns to be even more contrastive in most of the parameters regarding ageing voices. The acquisition study found that English rhythm takes longer to acquire than the syllable-timed rhythm of the French language, for instance (Grabe et al., 1999). Thus, the rhythmic prosodic properties gradually develop over the lifespan of an individual especially if a stress-timed, or accent-based speech rhythm is concerned.

Yet the study of how rhythmic prosodic properties change with age can mainly rely on the *apparent time* method of comparing the phonetic characteristics of speech produced by subjects who belong to different generations since a longitudinal study often remains an inaccessible task. The *apparent time* technique was used to study various age and gender peculiarities of American speech rhythm at the beginning of telephone conversations (Sokoreva & Shevchenko, 2016). The research results proved the idea that fundamental frequency is as important in identifying the rhythmic patterns of speech as duration: the variability of maximum pitch measures showed an increase with age that reach the highest point in the speech of the 'senior' group. This fact can be accounted for either by the lifespan development of an accent-based type of rhythm or by the instability of elders' speech (Linville, 2001). Thus, even the first phrases of a speaker in a telephone conversation can give you the key information about their social characteristics that are manifested by means of certain prosodic and eventually rhythmic features of speech.

Age-related changes in prosody typically described as a result of physical decline appear to be signs of speaking skills development when they are applied to middle age, a relatively advantageous stage of life, both socially and physically, before and at which most dramatic prosodic changes emerge. Through social practice people develop more effective ways and speech habits to optimise communication, and that experience has a direct impact on speech prosody. Thus, the adult life-long trajectory of prosody development proves to be gaining in effective communication, whatever the physical background may suggest.

In English language acquisition, age-related changes in prosody are described as development and improvement of language skills, while in adult life they are reported as features of decline and deterioration caused by biological processes (Beck, 2010; Helfrich, 1979; Linville, 2001). It was found that prosody development in American English, especially in middle-aged speakers, resulted in a sustained or enhanced contrast between accented and unaccented syllables by means of F0, intensity and duration due to the increase in accented syllables and feet durations, in a wider pitch range and in F0 variance. Despite the limitations of the telephone communication channel which diminishes the effects of pitch and intensity factors the contrast proved to increase with age (Shevchenko & Sokoreva, 2018). Physical decline and deterioration, previously reported in prosodic studies, appear to be counterbalanced by compensating changes which provide for accented syllables prominence that facilitates word recognition in speech.

Most of the changes occur before or at middle age, a life stage, which is argued to be positively assessed both socially and physically. Older people's speech remains intelligible and comprehensible due to their compensating ways of using prosody which facilitates intelligibility against the background losses in pitch and intensity maxima. The optimistic perspective on adult life-long prosody development gives evidence of the presence of age-related positive effects which facilitate communication.

'Age-related changes in prosody typically described as a result of physical decline appear to be signs of speaking skills development when they are applied to middle age, a relatively advantageous stage of life, both socially and physically, before and at which most dramatic prosodic changes emerge. Through social practice people develop more effective ways and speech habits to optimise communication, and that experience has a direct impact on speech prosody. Thus, the adult life-long trajectory of prosody development proves to be gaining in effective communication, whatever the physical background may suggest'

3.2. Developing methodology: a note on multimodality

The telephone corpus provides a unique opportunity to explore the effect of the human voice on the perception of the speaker's personality which is devoid of the visual channel information. Other media emphasise the role of visualisation in the modern world. In this context we could draw on the research which is aimed at comparing the relative force of the two types of information, aural and visual, in film making and Internet-mediated videos (Potapova et al., 2020). The perceptual range of interpretations of multimodal mono- and poly-ethnic communication was considered, and informational features perceived by the recipients were noted. Communication modalities were separated by the two tasks: first by listening to the recorded talks and then by watching the videos without the sound. Of particular interest for the present research is the ranking of features reported by the American participants in the experiment based on listening to the recorded material. Among the features which came on top of the list of prosodic clues for American speakers in contrast to the British ones were intensity and voice quality followed by pitch and tempo. Russian participants put tempo on top, while voice quality was also second (Potapova et al., 2020, p. 126). The data suggested that both intensity and voice quality are important parameters of speaker identification and assessment in American oral discourse, voice quality is likewise relevant for the perception of Russian listeners and should be included in our research agenda.

Further improvements in research methodology by shifting accents to certain parameters previously ignored could be also made on the basis of forensic phonetics findings. It was verified in recent research that not only maximum F0 values but also minimum F0 levels as well as mean pitch data present relatively reliable and constant points of reference in the pitch patterns employed by each individual (Potapova et al., 2022).

We, therefore, find it necessary to include and have a closer look at the parameters of both maximum and minimum F0 levels, as well as mean pitch, intensity, and voice quality features (jitter, shimmer, HNR) to observe age- and gender-related characteristics which could be part of personality identification and evaluation.

3.3. Intraspeaker variability

Although corpus analysis provides robust evidence of socially relevant demographic changes in the population which are bound to be reflected in their voices, there is another aspect of intra-speaker variability which is just as relevant for speaker identification. This is the knowledge of inter-speaker variability, part of which is conditioned by style, or genre of speech, the pragmatics of the speaker's intentions and their emotional state (Dunashova, 2021). The results of the recorded speech analysis, available in open sources through free access, proved to be Very specific, demonstrating an amazing diversity of images one great man, a linguist and lecturer, can put on. Concerning the general tendency of ageing voice losses, the 15 years period of time was found to operate against the sonority and resonance of the man's modal voice, but the lecturer managed to disguise it in his public performances (Dunashova, 2021).

The above-quoted study which is based on the speech production of an outstanding individual, nevertheless, calls for scientific scrutiny of case studies as well as big data research with special reference to age, gender and style as the major social factors affecting the voice of the individual. The study also confirmed the necessity to include voice quality characteristics which are prone to change with time.

4. STUDY RESULTS

4.1. Pitch characteristics

Aimed at revealing the variation in the prosodic forms of American English speakers and examining age-related and gender-specific prosodic features in telephone communication we analysed the speech samples of 30 American English citizens engaged in phone conversations. Studying the prosodic characteristics of three generations of people (young, middle-aged, senior) we obtain the human voice characteristics in the *apparent time*, i.e., in the dynamics of life-time development.

The findings of pitch values are summarised in Figure 1. The data obtained on the pitch of the interlocutors' voices vividly illustrates that middle-aged speakers, both male and female, have the highest *mean pitch* values. However, the interval between maximum and minimum pitch values (*pitch range*, or *pitch span*) is wider in the speech of senior people, both men and women.

These findings may be interpreted as the fact that middle-aged people have higher-pitched voices, but when they become older, they demonstrate a wider pitch range, i.e., a more skilful command of their voice. It may be explained by the fact that senior people try to convey their ideas more clearly, tend to influence younger interlocutors and be understood by them. Therefore, they use all the possibilities of voice, choosing the correct intonation with the help of high and low tones and highlighting important points.

Telephone communication also plays an important role here, since in the absence of visual contact it becomes harder for elderly people to effectively keep up the conversation with their interlocutors.

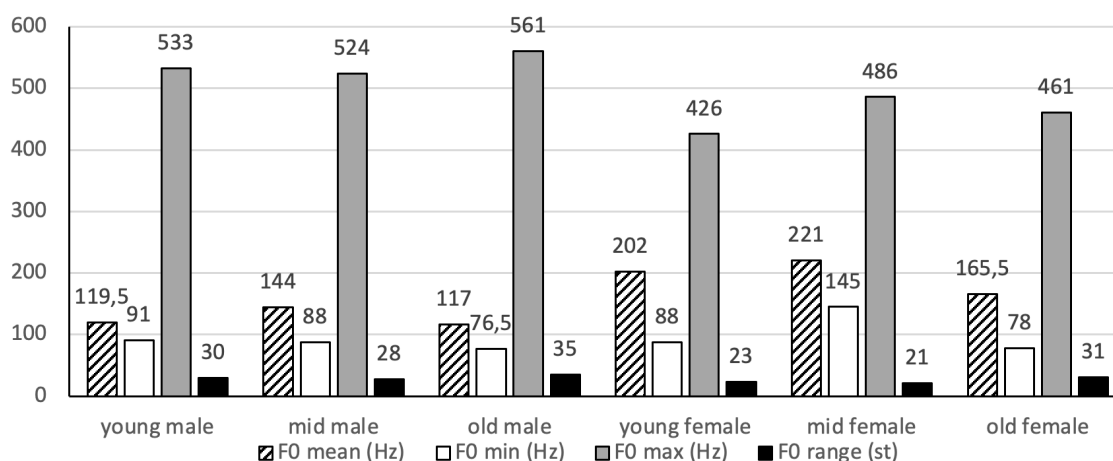


Figure 1. Pitch characteristics in three age groups across men and women

At the same time, it should be mentioned that the voice of senior people becomes lower with age, since both male and female senior speakers in the present research have the lowest *minimum pitch*. It may be accounted for by the tendency of the vocal folds to become less flexible with age and losing their ability to vibrate at a high pitch. It can be also pointed out that among men young participants have the highest minimum indicator, even higher than the minimum pitch of young female speakers. It is due to the fact that at a younger age men have a higher voice and are less able to produce low-pitched sounds.

As for women, their vocal folds develop differently. Female speakers in the present research corpus have the highest minimum pitch in middle age, which is much higher than that of men. However, in the voice of senior female interlocutors this figure then decreases sharply. That is why it can be said that at an older age women produce lower sounds. Factors that may account for this finding include *'intrinsic muscle weakening, ossification and calcification of laryngeal cartilages, and changes in vocal fold mass'* (Linville, 2001, p. 174).

As for *maximum pitch*, the data collected gives us grounds to suggest that men do not lose the ability to produce high pitches. Their pitch range widens, because along with lower and deeper sounds that appear with age, they still can produce higher-pitched sounds. Senior male speakers in the present research data have the highest maximum pitch levels in comparison with younger participants. This figure is even higher than that of all the female interlocutors, among whom middle-aged women have the highest maximum pitch. Therefore, it can be noted that older men possess higher-pitched voices, which may be due to physical changes of larynx. As a result of the disorganisation of collagen fibres and an increase in the density of the vocal folds male speakers are able to produce higher-pitched sounds, a tendency previously found in other studies (Linville, 2001, p. 172).

In senior women, on the contrary, the vocal folds atrophy and maximum pitch decreases. This is proven by the data of the present research which demonstrates that female participants

tend to reach their highest values of mean, minimum and maximum pitch in middle age but gradually lose that quality as they get older.

Thus, although both female and male speakers demonstrate the highest mean pitch in middle age, their pitch range widens as they get older. While men develop the ability to pronounce higher sounds with age, women's voices, in contrast, tend to phonate at lower levels. Men and women, as we can observe from the data retrieved, have different trajectories of pitch dynamics with age.

4.2. Intensity characteristics

In the process of *intensity* analysis, age-related differences have been reported both for men and women. The data obtained indicate that representatives of younger generations have the highest intensity (65 dB for male speakers and 66,5 dB for female speakers), although their values differ by only 3-5 decibels compared to scores of middle-aged participants (62 dB for men, 64 dB for women) and senior speakers (60 dB for men, 62 dB for women). This information may prove the fact that elderly men and women find it difficult to sustain high intensity during conversation. Lower intensity may appear as a result of age-related changes in the respiratory and phonatory system. Because of the loss of elasticity of lung tissues, senior speakers become less able to drive air out of the lungs with great force. Lower airway resistance in elderly men affects the valving capacity of their larynx and makes them speak quieter than young and middle-aged interlocutors. In addition, in the speech of senior men there are more glottal gaps that may also be a reason for lower intensity levels. For elderly women lower indicators may be due to the loss of respiratory and phonatory control with ageing.

4.3. Jitter and shimmer

Many scientists consider that coming to robust conclusions about the effect of ageing on *jitter* and *shimmer* levels is impossible (Linville, 2001, p. 174). Nevertheless, in the process of the

present analysis several results have been obtained. In telephone talks, it is senior speakers who have the highest jitter indicators (Figure 2). Both male and female senior people are ahead of younger representatives in this characteristic. Besides, male jitter values are much higher than female ones. It is difficult to find concrete factors that influence changes of vocal fold vibration, but elderly women exhibit greater aperiodicity in vocal fold vibrations which may be the reason for the increase in jitter indicators (Linville, 2001, p. 131). The present findings on jitter in female senior voices give more evidence on that point.

As for shimmer (Figure 3), the collected data does not show significant variation for women, while men's indicators differ only by 1-2%. Among male speakers the elderly people have the highest values.

Based on our data, it may be concluded that the results demonstrate exclusively considerable jitter increase as a function for older age respondents. It means that in the speech signal of senior speakers one period differs greatly from the next one. This is a sign that speech pathologies affecting the periodicity of the speech signal may develop with age, and female voices have proved to be healthier in that respect than male ones. However, the slight difference in shimmer of men's voices shows that speakers in the present study do not have any serious pathology which influences the amplitude perturbation of their voice signal. Consequently, although intensity values do not differ greatly in all ages, it can be said that in telephone communication young male and female speakers tend to speak louder in comparison to older interlocutors.

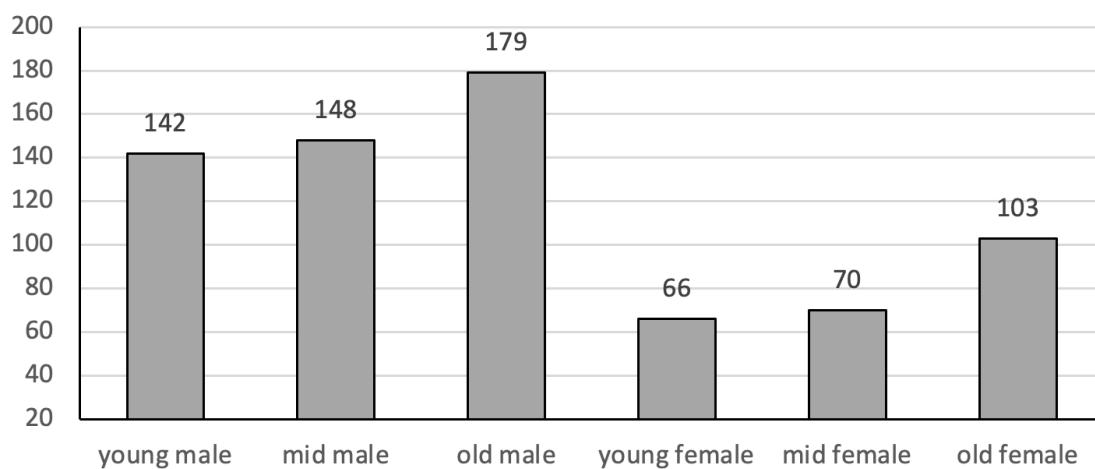


Figure 2. Jitter values (μs) in three age groups across men and women

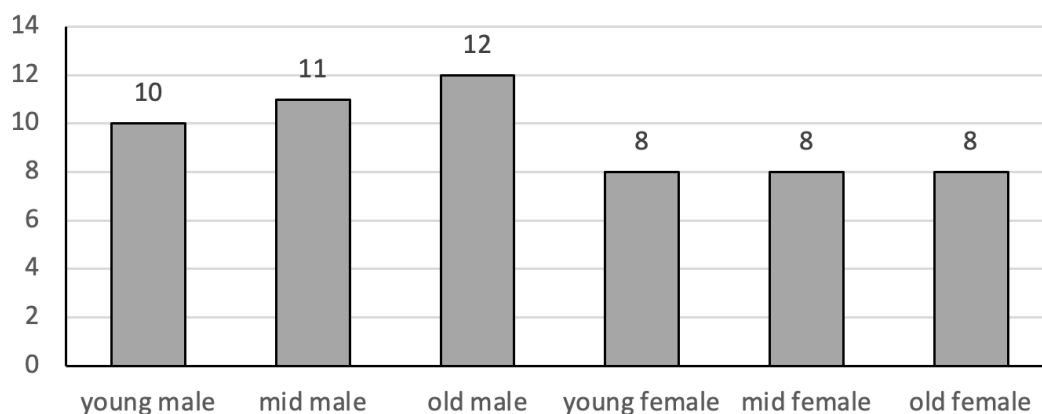


Figure 3. Shimmer (%) in three age groups across men and women

4.4. HNR characteristics

According to the analysis, the *harmonics-to-noise* ratio is almost the same for men of all ages (Figure 4). In the present research there is a small difference only in the voice of elderly

male speakers. However, this situation is quite unusual. The *harmonics-to-noise* ratio measures the amount of noise present in the voice signal. The lower the HNR is, the more noise, i.e., there is hoarseness in the speech. Since men's voices tend to become

hoarser with age, this indicator usually tends to be lower for them. But we can see this tendency in the speech of female interlocutors. The registered data shows that young participants have the highest HNR among women. It means that their speech is clearer, and their voices are more melodious. It is interesting to notice that young women's voices are much more harmonious than men's voices. Even the highest HNR among male speakers is almost twice as low compared to the female data. Nevertheless, the female HNR values decrease for elderly

women. Their voices become hoarser and breathier. HNR may also go down because of some laryngeal pathology such as incomplete closure of vocal folds or irregular glottal cycles.

There is a well-established tendency for younger people to possess more melodious voices in comparison to older generations. However, in the present study female speakers demonstrate this tendency in a more convincing way, while men's HNR is identical in young and middle-aged periods, with a slight drop in seniors.

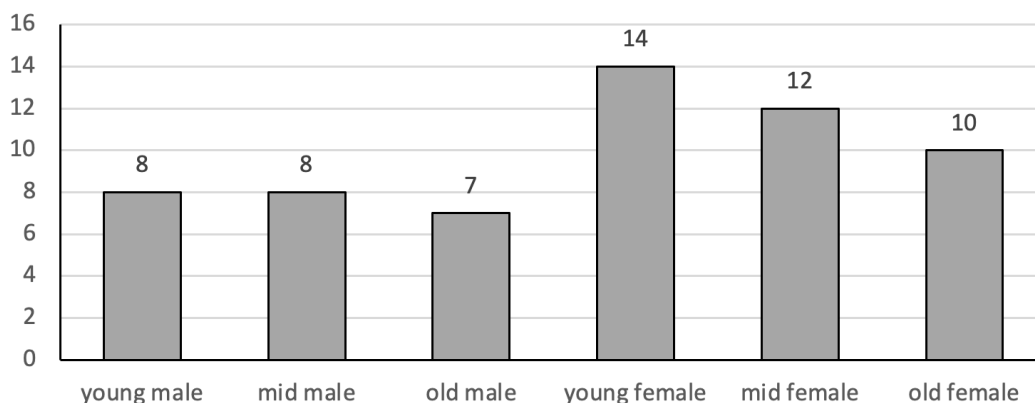


Figure 4. HNR (dB) in three age groups across men and women

5. CONCLUSION

Summarising all the data mentioned above, it can be concluded that the electronic analysis of telephone communication in English has revealed significant age-related phonetic changes in the speech of Americans which could be further specified as gender differences.

Firstly, among young, middle-aged, and senior speakers, both men and women of middle age have the highest mean pitch of voice. However, it is the elderly generation that demonstrates the widest pitch range. It has been discovered that senior men's voices acquire higher tones with age, while the voices of senior women become lower.

Secondly, although shimmer analysis has not shown substantial difference in all ages, the jitter level has turned out to be higher for senior speakers who have more periodic deviation in their voice signal.

Thirdly, in terms of intensity, we have found that young participants in telephone conversations have the highest indicators, i.e., they speak louder in comparison to older interlocutors.

Finally, data on the harmonics-to-noise ratio has displayed significant difference only in the speech of female representatives, among whom it is young women who have the most melodious and harmonious voices.

By comparing the data based on the measurements of eight parameters in three age groups and male and female gender we have found that although all the selected measurements proved to be relevant for either gender group, the age-related trajectory of voice dynamics may be different in male and female voices.

The common features are the dominance of young voices' intensity and HNR values, middle-aged voices' mean pitch and older people's pitch range at the expense of minimum pitch drop. A jitter increase with age is also common for both gender groups.

What makes the two gender groups different is the tendency for female mean pitch to be always higher, but for pitch maxima values to be still higher in the men's group, and what is especially noticeable, is the increase of pitch values in elderly men's voices, while elderly women lose their higher pitch prevalence.

There is also a difference in the degree of dominant voice quality change. Women's HNR drop is more impressive, whereas men's jitter increase with age shows higher values. There are physiological and cognitive reasons suggested by physiologists to account for the differences evidenced by the research.

Thus, gender- and age-specific prosodic features are some of the most important factors in identifying speakers in a telephone conversation. This study can be a sufficient basis for further research into professionally diverse types of telephone interaction between representatives from different social groups and can serve as a useful source of information about speech exchange on the phone for English learners.

By referring to the study of intraspeaker variability (see 3.4.), we can end on a positive note and present the results in a different light. We could illuminate the perspective of long-term lecturing or other forms of public speaking as well as singing which have proved to facilitate the preservation of clear voice quality and control of other vital speech characteristics.

References

- Beck, J. M. (2010). Organic variation of the vocal apparatus. In W. J. Hardcastle, J. Laver, & F. E. Gibbon, (Eds.), *The handbook of phonetic sciences* (pp. 155-201). Wiley-Blackwell.
- Boersma, P., & Weenink, D. (2017, November 12). Praat: Doing phonetics by computer. *University of Amsterdam*. <https://fon.hum.uva.nl/praat>
- Deliyski, D. D. (1993, September 22-25). Acoustic model and evaluation of pathological voice production. In *Proceedings of the 3rd European Conference on Speech Communication and Technology EUROSPEECH'93* (pp. 1969-1972). ISCA. <https://dx.doi.org/10.21437/Eurospeech.1993-445>
- Dunashova, A. (2021). Intraspeaker variability of a professional lecturer: Ageing, genre, pragmatics vs. voice acting (case study). In A. Karpov & R. Potapova (Eds.), *Speech and computer: Lecture notes in computer science* (pp. 179-189). Springer.
- Godfrey, J., Holliman, E., & McDaniel, J. (2021, November 15). SWITCHBOARD: Telephone speech corpus for research and development. *University of Pennsylvania*. <https://catalog.ldc.upenn.edu/products/LDC97S62>
- Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1201-1204). International Phonetics Association.
- Helfrich, H. (1979). Age markers in speech. In K. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 63-93). Cambridge University Press.
- Linville, S. E. (2001). *Vocal ageing*. Singular Publishing Group.
- Potapova, R., Potapov, V., & Kuryanova, I. (2022). Forensic identification of foreign-language speakers by the method of structural-melodic analysis of phonograms. In S. R. M. Prasanna, A. Karpov, K. Samudravijaya, & S. S. Agrawal (Eds.), *Speech and computer: Lecture Notes in Computer Science* (pp. 567-578). Springer. https://dx.doi.org/10.1007/978-3-031-20980-2_48
- Potapova, R. K., Potapov, V. V., & Komalova, L. R. (2020). *Perception of multimodal mono- and poly-ethnic communication*. Russian Academy of Sciences.
- Shevchenko, T., & Sokoreva, T. (2018). Corpus data on adult life-long trajectory of prosody development in American English, with special reference to middle age. In A. Karpov, O. Jokisch, & R. Potapova (Eds.), *Speech and computer: Lecture Notes in Computer Science* (pp. 606-614). Springer. https://dx.doi.org/10.1007/978-3-319-99579-3_62
- Shevchenko, T., & Sokoreva, T. (2019). First minute timing in American telephone talks: A cognitive approach. In A. Salah, A. Karpov, & R. Potapova (Eds.), *Speech and computer: Lecture Notes in Computer Science* (pp. 451-458). Springer. https://dx.doi.org/10.1007/978-3-030-26061-3_46
- Sokoreva, T., & Shevchenko, T. (2016). Starting a conversation: Indexical rhythmical features across age and gender (a corpus study). In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, speech, and dialogue: Lecture Notes in Computer Science* (pp. 495-505). Springer. https://dx.doi.org/10.1007/978-3-319-45510-5_57