AGE-RELATED VOWEL VARIATION IN FRENCH

Anne Hermes, Nicolas Audibert, Angélina Bourbon

Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, Paris
anne.hermes@sorbonne-nouvelle.fr, nicolas.audibert@sorbonne-nouvelle.fr, angelina.bourbon@sorbonne-nouvelle.fr

ABSTRACT

This paper presents a study on the temporal and spectral characteristics of oral and nasal vowels in French read speech, produced by speakers from 25 to 90 years. The study utilizes a set of eight vowel categories and combines classical vowel space-related metrics with MFCC parametrization to accurately include nasal vowels. In total 15,375 vowels were analyzed from data from 37 French speakers (20 females, 17 males). Our data showed (a) a slowing down (increase in vowel duration) with age and (b) no clear change in the overall centralization/dispersion of vowels with increasing age. However, for older male speakers, there was a decrease in the variability within vowel categories and a moderate increase in the separation between categories. Additionally, both male and female speakers demonstrated a greater acoustic distinction between oral and nasal vowels with aging.

Keywords: Aging, speech production, acoustics, French, MFCC, vowel space

1. INTRODUCTION

1.1. Age-affected speech parameters

Several studies investigated age-related effects on the supraglottal and glottal systems. Aging effects have been found in voice quality parameters (pitch: a decrease in F0 in women [1]; less clear for men [2, 3, 4]), in Voice Onset Time (shorter VOT in voiceless stops; [5, 6, 7], and in some instances, in formant frequencies (centralization, [8]). Overall, the most robust and documented age-related differences in speech production concern speech rate in that a slowing down is observed for older speakers [9, 10, 11, 12, 13, 14]. A recent study on French, [15] showed that older speakers produce longer pauses, longer sentence durations and lower articulation rates than younger speakers.

An important effect of aging over the lifespan (from young over middle-aged to older) has been reported by [16] for German and [17] for French in that effect, such as slowing down, and decrease in coarticulation is reported to be non-linear.

The relationships between aging and the realization of vowels have been addressed in various studies, focussing in most cases on the frequencies of the first two formants. While some authors report a decrease in formant frequencies, only the decrease in F1 for women appears to be rather robust across studies [18], but no clear tendency is found regarding vowel space area.

It is rather difficult to determine at what exact age these changes occur, as different studies have involved participants of different ages/age groups with different methods of analysis. Potential explanations for these changes in formant frequencies are inter alia the lengthening of the vocal tract [19], reduced and/or slowed down articulatory movement [10, 20], and other adjustments in articulation. Further research is needed to understand the relative contributions of these factors and to determine whether these changes are seen in all older adults or only in certain subgroups, such as those with health conditions. Normative data on vowel formant frequencies in older adults is necessary for interpreting acoustic changes that may occur in age-related conditions and for developing methods for automatically detecting speaker age.

To our best knowledge, such studies investigated individual formant frequencies and in some cases vowel space area, but did not examine the interplay between vowel categories and their evolution with aging. This study focuses on aging effects on temporal (vowel duration) and spectral shape in oral and nasal vowels in French read speech produced by speakers across different ages. This study aims to contribute to our knowledge of aging effects on speech production and better account for age-related vowel variation by using a rather large set of vowel categories (8) and combining classical vowel space-related metrics with MFCC parametrization to reliably include nasal vowels.

2. METHOD

2.1. Subjects

Acoustics recordings from 41 French native speakers were done in a quiet room with a head-mounted microphone (SHURE SM35-TQG), using Pyaudio (48 kHz sampling rate). The subjects' chronological age ranges from 23-90 years old. For the study, all of the subjects had normal vision or had corrected their vision to be normal, and none of them had any reported issues with speech, voice, language, psychological health, neurological functioning, or neurodegenerative disorders at the time of the study. Participants were screened for cognitive alterations using a standardised clinical tool, the Mini-Mental State Examination (MMSE, [21]). All participants were above the threshold of 24 (out of 30).

2.2. Speech material

The speech material is taken from Bourbon & Hermes (2021) [15] and consisted of French sentences (based on [22]), which differed in length and syntactic complexity (e.g., short-simple "Mélanie Dupont a réservé ses vacances" (Engl. *Mélanie Dupont booked her holidays*); long-complex condition "Mélanie Dupont a dit à sa mère, qui l'a appelée dans la matinée, qu'elle attend avec impatience ses vacances." (Engl. *Melanie Dupont told her mother, who called her in the morning, she's looking forward to her vacation*). Each sentence (short-simple, short-complex, long-simple, long-complex) was repeated three times, i.e., in total each speaker produced 36 sentences.

2.3. Annotations/Labelling

All sentences were segmented into phones using WebMaus [23]. After a sanity check, 2.2% of recordings with erroneous segmentation were discarded. To keep the number of utterances equal across speakers, two repetitions of each utterance were retained. We ended up with 37 speakers (20 females: x̄=62.5 years, sd=23.3, 17 males: x̄=56.7 years, sd=20.4). Fig. 1 displays density plots for the age distribution for female and male speakers.

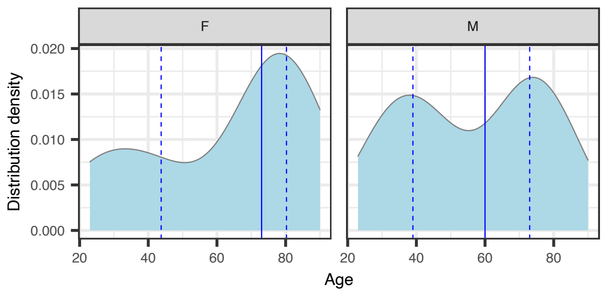

Figure 1: Density plots of age distribution in female (left) and male (right) speakers. The vertical lines represent the median (solid lines) and the quartiles (dashed lines).

The final corpus included exemplars of the oral vowels /i, y, e, ɛ, a, o/ and the nasal vowels /ɔ̃, ɑ̃/. After discarding the final vowel of each utterance, lengthened due to the accentuation pattern in standard French, for each vowel, we had a minimum of 14 exemplars (see Table 1).

Vowel	meanCount	minCount	maxCount
i	65	53	74
e	75	58	83
ɛ	51	40	56
a	118	96	122
o	35	26	36
y	16	14	18
ɔ̃	31	24	32
ɑ̃	25	18	27

Table 1: Exemplars for all oral and nasal vowels.

Note that we discarded /ɔ/ which was sparser and more appeared only in specific contexts. For the final dataset, we had a total of 15,375 vowels which were included in the analysis.

2.4. Measurements

For the temporal analysis, we extracted the durations of all target vowels. This temporal parameter was used as an indicator of speech rate and is expected to slow down with age as it has been reported in many studies before.

Automatic formants measurement is known to be error-prone, particularly in the case of nasalized segments [24]. Although thresholding may be applied to discard aberrant formant values [25], such a procedure may lead to discarding a large number of vowels (see e.g. [26]) and can hardly be applied to nasal vowels. In order to avoid including erroneous measures in analyses, we use MFCC measures to reliably capture the spectral characteristics of target vowels, following [26]. A custom Praat [26] script is applied to extract on each target vowel 12 MFCCs (not including coefficient 0 related to the overall level of energy) on a 15 ms frame centred on the middle of the vowel, using a filter bank spaced by 100 Mel.

Separately for each speaker, different metrics *DistCentroid*, *VDispersion* and *ContrastLoss* [28] are computed from 12-dimensional MFCC values, resulting in one value for each metric for each vowel exemplar. *DistCentroid* is defined as the Euclidean distance from a vowel exemplar to the vowel space centroid (computed as the centroid of vowel-specific centroids); *VDispersion* as the Euclidean distance from an exemplar to the centroid of the corresponding vowel category; *ContrastLoss* as the posterior probability for the exemplar to belong to the reference vowel category, computed by a linear discriminant analysis (LDA). In addition to ContrastLoss, confusion rates between vowel categories were also extracted from the LDA results. The R code for the computation of all metrics applied is provided as a Shiny interactive application and for download under https://shiny.laboratoirephonetiquephonologie.fr/vowel_space_metrics_computation/. Since we assume that there are differences due to age, we display in addition to the results for all speakers, the results for the older speakers separately (≥ 60 years; solid lines).

3. RESULTS

3.1. Age effects on vowel duration

For the analysis of this dataset, we analyzed male and female speakers separately, since it is known that age-related effects are reported to be different.

Fig. 2 displays the MFCC for all vowel durations across male and female speakers over age. In line with what has been reported for age-related effects on temporal parameters, our data also revealed a slowing down with increasing age (females: r=0.554; males: r=0.789). Comparable trends are observed for the older speakers, with a larger between-speakers variability (females: r=0.204; males: r=0.385).

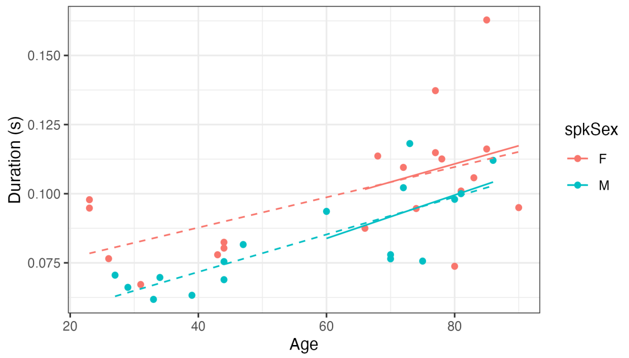

Figure 2: Scatterplot of mean duration per speaker against age, for female and male speakers. The regression is illustrated by dashed lines for all speakers, and by solid lines for older speakers.

3.2. Age effects on formants

As said before, for the formant analysis, we used several metrics to look at age-related effects. For the metric *DistCentroid* (vowel exemplar to the vowel space centroid), no overall pattern of change with age is found either for female (r=.106) or male speakers (r=.083), as shown by Fig. 3.

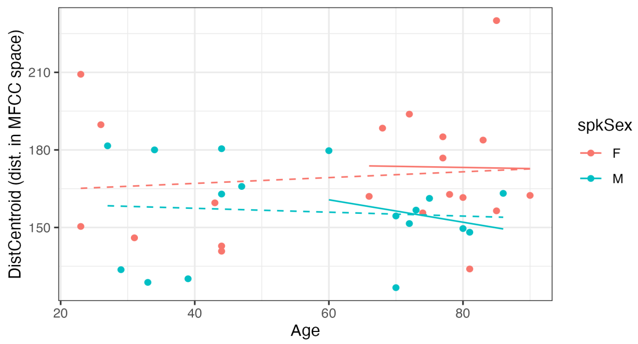

Figure 3: Scatterplot of mean DistCentroid values per speaker against age, for female and male speakers. Higher values of DistCentroid correspond to a more dispersed vowel space. The regression is illustrated by dashed lines for all speakers, and by solid lines for the older speakers.

A moderate tendency toward centralization is found for older male speakers (r=-.233) but not for older female speakers (r=-.013).

For the metric *VDispersion* (distance from exemplar to the centroid of the corresponding vowel category), Fig. 4 illustrates diverging patterns found for female (r=.211) and male speakers (r=-.387) across the lifespan. For older speakers, category-wise variability decreases with age for male speakers (r=-.569) while this tendency is much weaker for female speakers (r=-.120).

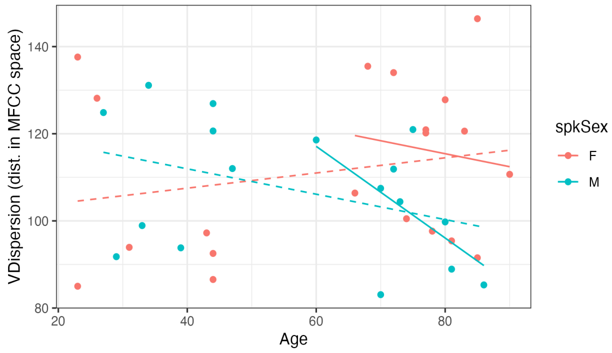

Figure 4: Scatterplot of mean VDispersion values per speaker against age, for female and male speakers. Higher values of VDispersion correspond to a larger variability within vowel categories. The regression is illustrated by dashed lines for all speakers, and by solid lines for the older speakers.

For the metric *ContrastLoss* (posterior probability of vowel category membership), Fig. 5 illustrates that there is no effect of age on the metric for female speakers (r=.035), neither across ages in the older group (r=.085). For male speakers (r=.343), it can be observed that overlap decreases with age, across all ages and in the older group specifically (r=.300). Note that *ContrastLoss* analysis considers all vowel categories together and may hide category-specific patterns of change with aging.

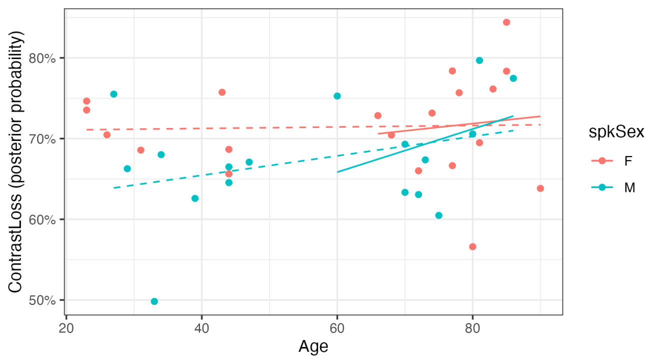

Figure 5: Scatterplot of mean ContrastLoss values per speaker against age, for female and male speakers. Higher values of ContrastLoss correspond to a lesser overlap between vowel categories. The regression is illustrated by dashed lines for all speakers, and by solid lines for the older speakers.

In order to get more insights into how the organization of the speaker's vocalic system is affected by aging, the evolution with age of confusion attributed by the LDA between broader vowel classes was further explored. Although diverging patterns are observed for most confusion types considered, a moderate but consistent tendency is found for confusions between oral and nasal vowels, which undergo less acoustic overlap with aging, as illustrated in Fig. 6 (females: r=-.533; males: r=-.235). However, the lesser overlap between oral and nasal vowels is reached earlier by female speakers, which may explain the diverging patterns observed between female and male speakers in the older group (females: r=.244; males: r=-.47).

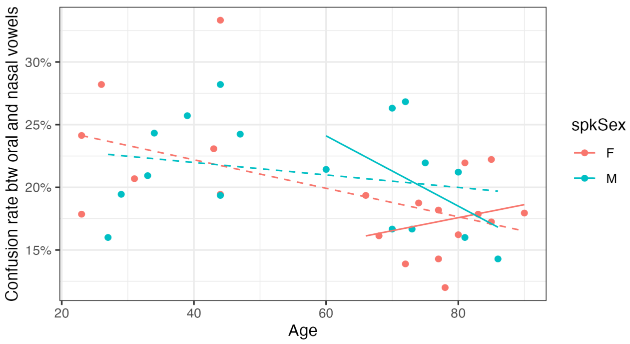

Figure 6: Oral-nasal confusions based on LDA classification per speaker against age, for female and male speakers. The regression is illustrated by dashed lines for all speakers, and by solid lines for the older speakers.

4. DISCUSSION AND CONCLUSION

In this study, the effects of age on vowel duration and formants were investigated using a dataset of French read speech produced by male and female speakers of different ages.

In line with what has been reported in the literature on aging and speech, the results also showed that speech, here measured as vowel duration, is slowed down with increasing age for both male and female speakers. For formants, no overall pattern of change with age was found for either male or female speakers. However, a moderate tendency toward centralization was found for older male speakers, and there were diverging patterns of change with age for female and male speakers in terms of vowel category variability and overlap between vowel categories. Specifically, category-wise variability decreased with age for male speakers, while this tendency was much weaker for female speakers. There was also a moderate but consistent tendency for acoustic overlap between oral and nasal vowels to decrease with age, with this reduction in overlap being reached earlier by female speakers. These findings suggest that aging may differentially affect the production of vowel sounds in men and women. They also suggest that in languages such as French in which the contrast between oral and nasal vowels is phonological, the acoustic separation between those categories may increase with aging. Further research is needed to understand the underlying mechanisms and potential consequences of these changes.

5. ACKNOWLEDGEMENTS

This work was supported by the French ANR ANR-10-LABX-0083 (Labex EFL) and ANR-21-CE28-0015-01 (PASDCODE) and by the Swiss NSF Sinergia Program (ChaSpeePro, CRSII5_202228).

6. REFERENCES

[1] L. A. Ramig and R. L. Ringel, "Effects of Physiological Aging on Selected Acoustic Characteristics of Voice," *Journal of Speech, Language, and Hearing Research*, vol. 26, no. 1, pp. 22–30, Mar. 1983, doi: 10.1044/jshr.2601.22.

[2] S. E. Linville, "The sound of senescence," *Journal of Voice*, vol. 10, no. 2, pp. 190–200, Jan. 1996, doi: 10.1016/s0892-1997(96)80046-4.

[3] R. J. Baken, "The Aged Voice: A New Hypothesis," *Journal of Voice*, vol. 19, no. 3, pp. 317–325, Sep. 2005, doi: 10.1016/j.jvoice.2004.07.005.

[4] U. Reubold, J. Harrington, and F. Kleber, "Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers," *Speech Communication*, vol. 52, no. 7–8, pp. 638–651, Jul. 2010, doi: 10.1016/j.specom.2010.02.012.

[5] B. J. Benjamin, "Phonological performance in gerontological speech," *Journal of Psycholinguistic Research*, vol. 11, no. 2, pp. 159–167, Mar. 1982, doi: 10.1007/bf01068218.

[6] R. J. Morris and W. S. Brown, "Age-related differences in speech variability among women," *Journal of Communication Disorders*, vol. 27, no. 1, pp. 49–64, May 1994, doi: 10.1016/0021-9924(94)90010-8.

[7] J. Ryalls, A. Cliche, J. Fortier-blanc, I. Coulombe, and A. Prud'hommeaux, "Voice-onset time in younger and older French-speaking Canadians," *Clinical Linguistics & Phonetics*, vol. 11, no. 3, pp. 205–212, Jan. 1997, doi: 10.3109/02699209708985191.

[8] M. P. Rastatter, R. A. McGuire, J. Kalinowski, and A. Stuart, "Formant Frequency Characteristics of Elderly Speakers in Contextual Speech," *Folia Phoniatrica et Logopaedica*, vol. 49, no. 1, pp. 1–8, 1997, doi: 10.1159/000266431.

[9] M. Bilodeau-Mercure, V. Kirouac, N. Langlois, C. Ouellet, I. Gasse, and P. Tremblay, "Movement sequencing in normal aging: speech, oro-facial, and finger movements," *AGE*, vol. 37, no. 4, Jul. 2015, doi: 10.1007/s11357-015-9813-x.

[10] A. Hermes, J. Mertens and D. Mücke, "Age-related effec ts on sensorimotor control of speech production", *Interspeech*, 2018, pp. 1526-1530.

[11] B. L. Smith, J. Wasowicz, and J. Preston, “Temporal Characteristics of the Speech of Normal Elderly Adults,” *Journal of Speech, Language, and Hearing Research*, vol. 30, no. 4, pp. 522–529, Dec. 1987, doi: 10.1044/jshr.3004.522.
[12] E. D. Mysak and T. D. Hanley, “Aging Processes in Speech: Pitch and Duration Characteristics,” *Journal of Gerontology*, vol. 13, no. 3, pp. 309–313, Jul. 1958, doi: 10.1093/geronj/13.3.309.
[13] P. H. Ptacek, E. K. Sander, W. H. Maloney, and C. C. R. Jackson, “Phonatory and Related Changes with Advanced Age,” *Journal of Speech and Hearing Research*, vol. 9, no. 3, pp. 353–360, Sep. 1966, doi: 10.1044/jshr.0903.353.
[14] J. D. Amerman and M. M. Parnell, “Speech timing strategies in elderly adults,” *Journal of Phonetics*, vol. 20, no. 1, pp. 65–76, Jan. 1992, doi: 10.1016/s0095-4470(19)30254-2.
[15] A. Bourbon and A. Hermes, “Have a break: Aging effects on sentence production and structuring in French”, *International Seminar on Speech Production,* 2021, pp.102-105.
[16] T. Thies, A. Hermes, and D. Mücke, “Compensation in Time and Space: Prominence Marking in Aging and Disease,” *Languages*, vol. 7, no. 1, p. 21, Jan. 2022, doi: 10.3390/languages7010021.
[17] D. D’Alessandro and C. Fougeron, “Changes in Anticipatory VtoV Coarticulation in French during Adulthood,” *Languages*, vol. 6, no. 4, p. 181, Oct. 2021, doi: 10.3390/languages6040181.
[18] R. D. Kent and H. K. Vorperian, “Static measurements of vowel formant frequencies and bandwidths: A review,” *Journal of Communication Disorders*, vol. 74, pp. 74–97, Jul. 2018, doi: 10.1016/j.jcomdis.2018.05.004.
[19] S. E. Linville and J. Rens, “Vocal Tract Resonance Analysis of Aging Voice Using Long-Term Average Spectra,” *Journal of Voice*, vol. 15, no. 3, pp. 323–330, Sep. 2001, doi: 10.1016/s0892-1997(01)00034-0.
[20] P.J. Watson and B. Munson. "Parkinson’s disease and the effect of lexical factors on vowel articulation", *The Journal of the Acoustical Society of America* 124, no. 5, 2008, EL291-EL295.
[21] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “‘Mini-mental state,’” *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, Nov. 1975, doi: 10.1016/0022-3956(75)90026-6.
[22] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, “Acoustic and respiratory evidence for utterance planning in German,” *Journal of Phonetics*, vol. 41, no. 1, pp. 29–47, Jan. 2013, doi: 10.1016/j.wocn.2012.08.007.
[23] T. Kisler, U. Reichel, and F. Schiel, “Multilingual processing of speech via web services,” *Computer Speech & Language*, vol. 45, pp. 326–347, Sep. 2017, doi: 10.1016/j.csl.2017.01.005.
[24] F. de Wet, K. Weber, L. Boves, B. Cranen, S. Bengio, and H. Bourlard, “Evaluation of formant-like features on an automatic vowel classification task,” *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1781–1792, Sep. 2004, doi: 10.1121/1.1781620.
[25] C. Gendrot and M. Adda-Decker, “Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German,” *Interspeech*, Nov. 2005, pp. 2453-2456.
[26] E. Ferragne and F. Pellegrino, “Vowel systems and accent similarity in the British Isles: Exploiting multidimensional acoustic distances in phonetics,” *Journal of Phonetics*, vol. 38, no. 4, pp. 526–539, Oct. 2010, doi: 10.1016/j.wocn.2010.07.002.
[27] P. Boersma and D. Wennink, “Praat: doing phonetics by computer [Computer program]”, Version 6.3.03, retrieved 17 December 2022 from http://www.praat.org/.
[28] N. Audibert, C. Fougeron, C. Gendrot and M. Adda-Decker, "Duration-vs. style-dependent vowel variation: A multiparametric investigation", 1*8th International Congress of Phonetic Sciences*, 2015, pp. 5-10.