



# How the language we speak determines the transmission of COVID-19

Georgios P. Georgiou<sup>1,2</sup> · Chris Georgiou<sup>3</sup> · Ahmad Kilani<sup>4</sup>

Received: 6 December 2020 / Accepted: 25 December 2020  
© Royal Academy of Medicine in Ireland 2021

## Abstract

**Background** Little body of research has focused on the epidemic transmissibility and language interface.

**Aims** In this paper, we aim to investigate whether (i) the feature of aspiration found in the phonological inventory of several languages and (ii) the frequency of occurrence of stop consonants are associated with the transmission of COVID-19 among humans.

**Methods** The study's protocol was based on a corpus of countries infected by COVID-19 and of which the linguistic repertoire includes a widely spoken language in individuals' everyday communication. We tested whether languages with and without aspiration differ in terms of COVID-19 reproduction number, and whether the frequency of occurrence of stop consonants in several languages correlates with the virus reproduction number.

**Results** The results demonstrated no significant effect of aspiration on the transmission of the virus, while a positive correlation between the frequency of occurrence and transmissibility was observed only for the consonant /p/; this might suggest that languages that use /p/ more frequently might spread the virus more easily.

**Conclusions** The findings of this study can offer a tentative picture of how speaking specific sounds can be associated with COVID-19 transmissibility.

**Keywords** Aspiration · Consonants · COVID-19 · Transmission

## Introduction

COVID-19 triggered by SARS-CoV-2 infection has resulted in a global pandemic in 217 nations with more than 42.6 million infections and 1.1 million deaths as of 24 October 2020, according to the COVID-19 Map of the Johns Hopkins Coronavirus Resource Center. Currently, there is no specific drug or vaccine against COVID-19 [1]; therefore, effective preventing measures will play a vital role in controlling the spread of the virus. Nonetheless, epidemiologic data

demonstrate a rampant increase of cases in some countries, while others seem to have flattened the number of cases. This advances the question of whether a holistic scientific apprehension of disease spreading channels has yet to be achieved, and consequently whether there are more effective methods to prevent its transmission.

The most important characteristic of COVID-19 is that it can easily be transmitted from human to human. Infectious microorganisms in the air released from infected individuals through expiratory activities and the inhalation of these microorganisms by other individuals result in the transmission of the disease [2, 3]. So, COVID-19 can be transmitted through sneezing or coughing due to the high velocity of droplets produced by these activities. However, recent studies have demonstrated that large quantity of droplets may be produced even by talking or breathing [4, 5]. According to Lindsley et al. [4], droplets produced through breathing are more than those produced by coughing since the latter takes places with lower frequency in comparison to breathing. Considering evidence from the World Health Organization, one-third of individuals infected by COVID-19 do not have cough, and therefore the virus is

✉ Georgios P. Georgiou  
georgiou.georg@unic.ac.cy

<sup>1</sup> Department of Languages and Literature, University of Nicosia, Nicosia, Cyprus

<sup>2</sup> Department of General and Russian Linguistics, RUDN University, Moscow, Russia

<sup>3</sup> Specialist in Computer Programming, Lamaca, Cyprus

<sup>4</sup> Department of Comparative Politics, RUDN University, Moscow, Russia

more possible to be transmitted through the emission of aerosol particles from talking or breathing.

The transmissibility or contagiousness of infectious diseases such as COVID-19 can be estimated using the basic reproduction number ( $R_0$ ) [6]. This metric is not constant since it relies on the duration of the infectious period, the likelihood of infecting an individual during one contact, and the number of new individuals that an infected individual has contacted per unit of time [7].  $R_0$  varies not only from disease to disease but also for the same disease across different populations.

Some studies assume that certain types of sounds used in world languages produce more droplets compared to other sounds; this might have significant impact for the transmission of viruses according to the language we speak. For example, Asadi et al. [8] investigated the effect of voicing and articulation manner on aerosol particle emission during human speech. The authors measured the particle emission rates of 56 healthy individuals who produced phones in isolation and spoken speech. The results showed that some vowels (e.g., /i/) produce more particles than others (e.g., /a/), and voiced plosive consonants (e.g., /b/) produce more particles than voiceless fricatives (e.g., /f/). Abkarian and Stone [9] provide novel evidence about the mechanisms that create droplets in the mouth. The researchers recorded a high-speed video of a volunteer who produced various sounds. The findings demonstrated that the consonants /b d p t/ created the most saliva because they involve a burst of air through a narrow saliva-filled space. By contrast, consonants such as /m/ produce only a few droplets because the air is sent through the nose. All consonants which were found to create a lot of droplets during speech have the same manner of articulation; they are stop or plosive consonants. Such consonants are produced with a complete closure of the articulators (e.g., lips, tongue) which impedes the air from escaping the mouth. When the articulators separate with each other the air is released in a small burst of sound [10].

Inouye [11] developed a controversial hypothesis to justify the fact that Japanese tourists in China in 2003 were not infected by SARS in contrast to American tourists. The author proposes that the use of aspirated consonants increases the chance for the transmission of SARS from human to human since such consonants emit a lot of droplets compared to other types of sounds. He emphasizes the possibility that Chinese shop assistants were speaking to Japanese tourists in Japanese, a language in which aspiration is weak, while they were speaking to American tourists in English, which has stronger aspiration; this might explain the zero infection of Japanese tourists. A follow-up study of Inouye and Sugihara [12] provided further support to this hypothesis, showing that the pressure of wind and the strength of puff are weaker for the Japanese language in comparison to English and Chinese. Aspiration is a period of voicelessness after the articulation

of a stop consonant and prior to the beginning of the vowel voicing [10]. Thus, when aspirated consonants are produced, a burst of air comes out of the mouth.

Similarly to Inouye [11], Georgiou and Kilani [13] devised the hypothesis that aspirated consonants might increase the transmission of COVID-19. The authors compared the number of COVID-19 cases per million of population in 26 countries which were mostly infected by the virus. They divided the countries into two lists: countries of which the dominant language contains aspirated consonants and countries of which the dominant language does not contain aspirated consonants. It was observed that countries with languages that include aspiration had more cases of COVID-19; however, there were no significant differences in the number of cases between the two types of languages. Still, any conclusions would be controversial and uncertain due to methodological limitations.

In this study, we aim to provide more detailed insights into how the production of particular consonants during speech might contribute to the spread of COVID-19. To our knowledge, such investigations are extremely limited in the literature—we aim to highlight a novel relationship between linguistics and biological sciences. We rest upon two hypotheses. First, we assume that the use of aspirated consonants during speech might relate to the transmission of COVID-19. This is because aspiration involves a puff of air and, subsequently, more droplets might be emitted from the mouth compared to non-aspirated productions. Second, we assume that the frequency of occurrence of specific consonants that are said to produce a lot of droplets during speech (see [9]) will positively correlate with the transmissibility of COVID-19, which can be reflected in the  $R_0$  of the disease.

## Analysis 1

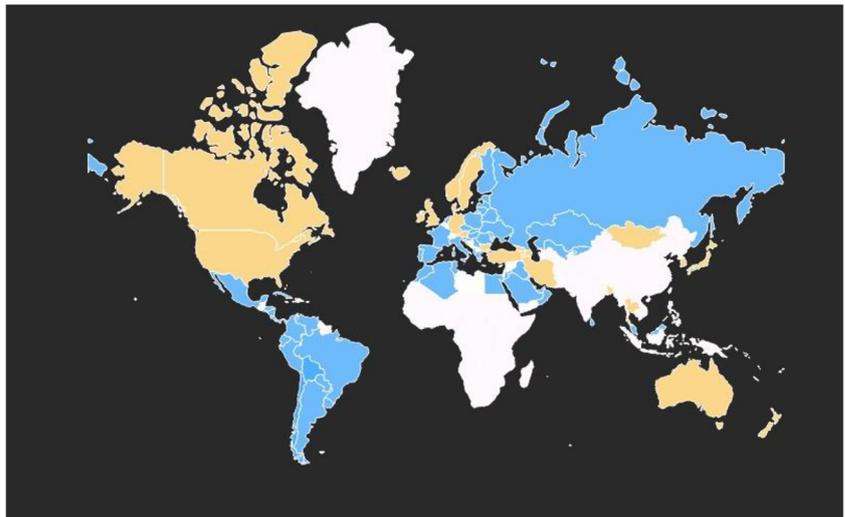
This analysis aimed at investigating whether languages with aspirated consonants would transmit easier COVID-19 in comparison to languages without aspirated consonants.

## Methodology

### Sample

We initially have chosen the 150 most infected countries by COVID-19 as of October 17, 2020 [14]. A country does not represent a particular language, so, in order to control as much as possible this factor, we selected countries of which approximately three-quarters of the population uses a particular standard language for everyday communication; all the other countries that did not meet this criterion were excluded (e.g., Switzerland, Cameroon, Myanmar). The number of selected countries was 91.

**Fig. 1** List of the countries used in the analysis. Blue color represents countries of which the primary language does not have aspirated consonants and yellow color represents languages with aspirated consonants.



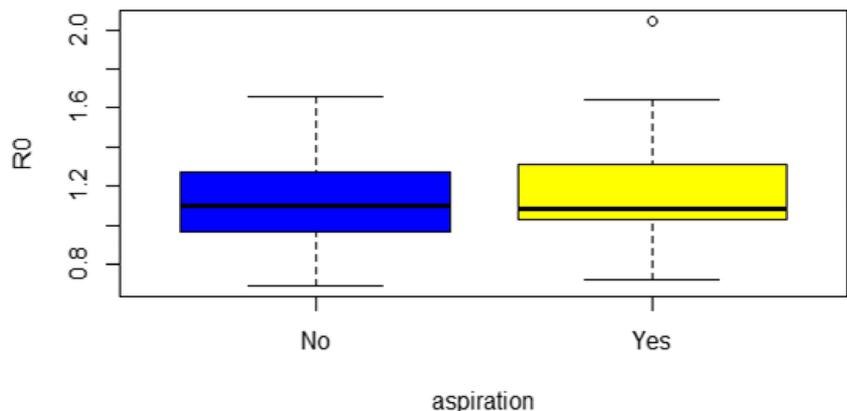
We also gathered information about the  $R_0$  of COVID-19 for the countries in our list as of October 17, 2020. This information was retrieved from the Epidemic Forecasting: COVID-19 (<http://epidemicforecasting.org>) which is managed by the Future of Humanity Institute, University of Oxford. We could not retrieve information for 8 countries, and thus the final number of the countries included in our database was 83 (with aspiration:  $n = 25$ , without aspiration  $n = 58$ ) (see Fig. 1).

To collect information about the existence of aspirated consonants in the phonological inventory of standard languages, we used PHOIBLE [15], a database that contains 3020 inventories from 2186 languages drawn from several sources such as UPSID [16], South American Phonological Inventory Database [17], the Stanford Phonology Archive [18], and other secondary sources (see [19]).

### Statistical Analysis

Our protocol was based on a point-biserial correlation test conducted in R [20]. This kind of test was the most appropriate since we had a dichotomous independent variable, *Aspiration* (Yes/No), and a scale dependent variable,  $R_0$ .

**Fig. 2** Boxplot of the point-biserial correlation test



### Results

The results of the statistical analysis showed that languages with aspirated consonants had higher  $R_0$  ( $M = 1.19$ ) than languages without aspirated consonants ( $M = 1.14$ ) (see Fig. 2). Nevertheless, these differences were not significant (see Table 1 for the results of the statistical model).

### Analysis 2

The second analysis aimed at investigating the correlation between the frequency of occurrence of four consonants found in particular languages and the  $R_0$  of COVID-19 in the countries in which these languages are primarily spoken.

### Methodology

#### Sample

The sample of the analysis consisted of the frequency of occurrence of the consonants /b d p t/ found in 16 languages,

**Table 1** Results of the point-biserial correlation test

data: test\$ro and test\$aspiration
$t = 0.84638$ , $df = 81$ , $p$ value = 0.3998
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
- 0.1245758 0.3031956
sample estimates:
cor
0.09362932

which are mainly spoken in 16 different countries. The data was retrieved from Peust [21], which includes a corpus of the frequency of occurrence of phonemes found in 50 languages; the corpus was developed out of analyses of spoken and written speech from 10,000 to 150,000 words. For our analysis, we only selected languages that are represented in our initial list. We selected the consonants /b d p t/ as there is recent evidence that they can produce a lot of droplets during speech [9]. Note that these consonants are very common in world languages with their frequency to span from 60 to 80% [15].

### Statistical analysis

We used a series of correlation tests in R. The first variable was the frequency of occurrence for each of the four consonants in particular languages (counted in percentages), and the second variable was the  $R_0$  of COVID-19 in the countries where these languages are mostly spoken (as of 17 October 2020).

### Results

The results of the statistical analysis showed a negligible or small negative correlation between  $R_0$  and /b/ ( $r(14) = -0.23$ ,  $p > .05$ ), /d/ ( $r(14) = -0.03$ ,  $p > .05$ ), and /t/ ( $r(14) = -0.02$ ,  $p > .05$ ), and a small positive correlation between  $R_0$  and /p/ ( $r(14) = 0.13$ ,  $p > .05$ ). Figure 3 illustrates the results of the analysis.

We conducted another correlation test using data for consonant frequencies in several languages. This data was taken from an online source<sup>1</sup>, which includes the frequency of occurrence of letters found in texts; the texts were collected from various sources. Not all the languages in this data were the same as those in the previous data. The analysis showed a negative correlation for /b d t/, but a large positive correlation for /p/ ( $r(12) = 0.55$ ,  $p < .05$ ) (see Fig. 4); this would

practically mean that languages that use /p/ more frequently have more chance to spread the virus.

### Discussion

We conducted two analyses to determine the relationship between the transmissibility of COVID-19 and the use of aspirated consonants, and the relationship between the transmissibility of COVID-19 and the frequency of occurrence of specific consonants in several languages.

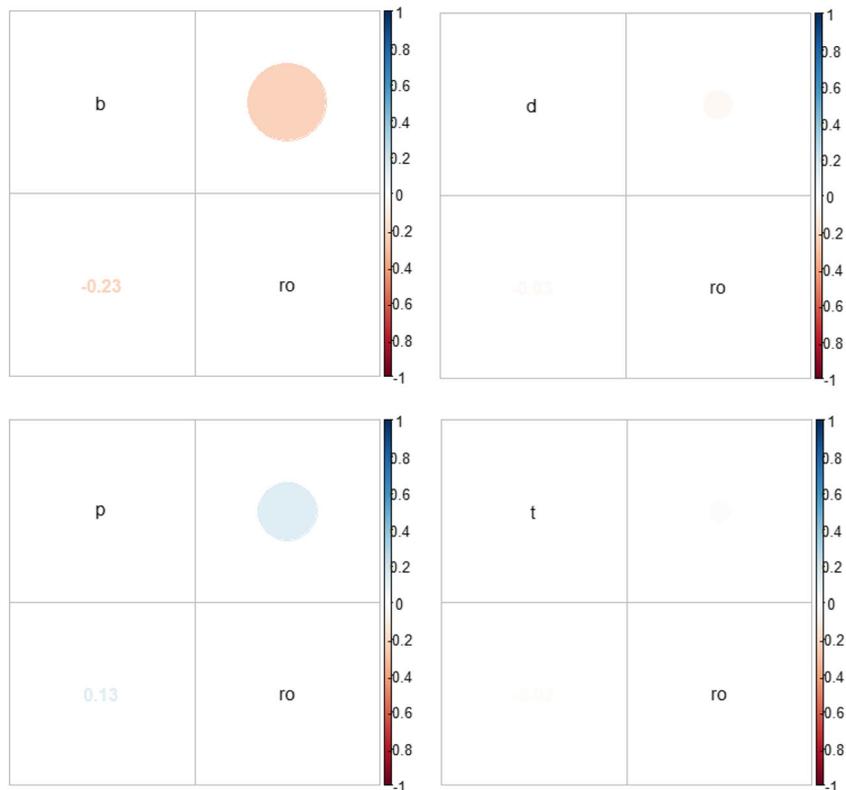
The findings portrayed that there were no significant differences for the transmissibility of the virus between countries that mainly use a particular language that contains aspirated consonants and countries with languages that do not contain aspirated consonants. This corroborates the earlier findings of Georgiou and Kilani [13] who found no significant differences between the group of languages that use and the group of languages that do not use aspirated consonants. The commonplace between the two studies is that there were more cases of COVID-19 and the virus was more transmittable in languages that include aspirated consonants. However, the statistical analysis indicated that this difference was not important and therefore our initial hypothesis cannot be accepted.

The results of the second analysis showed that the frequency of occurrence of /b d t/ did not have any positive correlation with the transmission of the disease. There was a small correlation for /p/ and upon conducting another analysis with different data, we found a large correlation. According to Abkarian and Stone [9], who investigated stop consonant saliva productions, although both /b/ and /p/ consonants produce a lot of droplets during speech, /p/ surpasses /b/ in terms of droplet emission; this took place when the speaker produced the “Ba-aBa-aB” and “Pa-aPa-aP” sequences. The aforementioned findings can be explained from the fact that /b/ is a voiced consonant, and thus there is vibration of the vocal folds, which leads to rapid pressure modulations in the air-flow. In that way, filaments are destabilized by these quick movements resulting in the production of fewer droplets for /b/. So, this might explain the positive correlation of /p/ with the virus transmissibility, suggesting that languages with more frequent use of the /p/ sound may have more chances to spread the virus.

This study offers only a tentative picture of how the use of consonants is associated with the transmissibility of COVID-19. The conclusions drawn from the results cannot be generalized at the moment due to methodological limitations. First, we were not able to include a large number of languages, considering that there are more than 7000 active languages in the world. Second, we do not know the exact linguistic background of each individual and which languages they use during their everyday communication. For example, one might be a native speaker of the predominant language of a

<sup>1</sup> [https://en.wikipedia.org/wiki/Letter\\_frequency](https://en.wikipedia.org/wiki/Letter_frequency)

**Fig. 3** The results of the correlation analysis



country, but they may use another language for communication when speaking with other individuals. Third, the transmissibility of the virus may also depend on other factors such as social distance measures and other sources of transmission (e.g., coughing, contact with an infected surface). Although it is difficult to determine the exact relationship between the transmission of a virus and the use of language, future studies

may rely on this study to perform further research in controlled environments.

**Data availability** Not applicable

**Compliance with ethical standards**

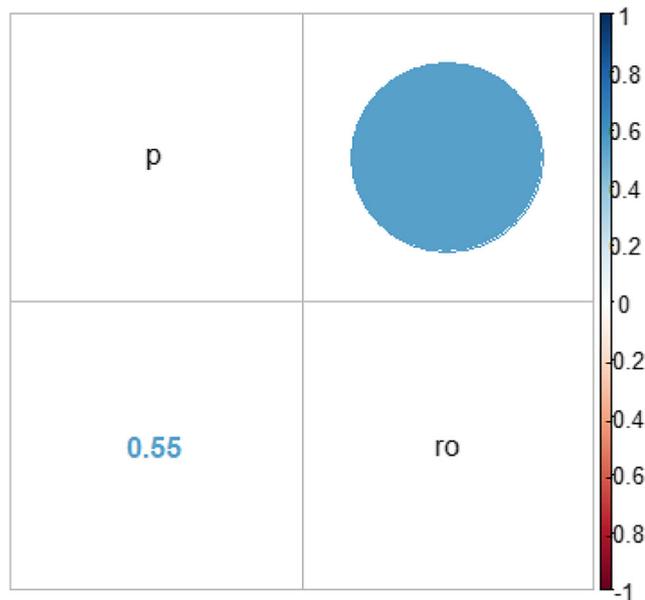
**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** Not applicable

**Consent to participate** Not applicable

**Consent for publication** Not applicable

**Code availability** Not applicable



**Fig. 4** The results of the second correlation analysis for /p/

**References**

1. Newton PN, Bond KC, Adeyeye M et al (2020) COVID-19 and risks to the supply and quality of tests, drugs, and vaccines. *Lancet Glob Health* 8:e754–e755
2. Tellier R (2006) Review of aerosol transmission of influenza A virus. *Emerg Infect Dis* 12(11):1657–1662
3. Weber TP, Stilianakis NI (2008) Inactivation of influenza A viruses in the environment and modes of transmission: a critical review. *J Infect* 57(5):361–373

4. Lindsley WG, Blachere FM, Beezhold DH et al (2016) Viable influenza A virus in airborne particles expelled during coughs versus exhalations. *Influenza Other Respir Viruses* 10(5):404–413
5. Yan J, Grantham M, Pantelic J et al (2018) Infectious virus in exhaled breath of symptomatic seasonal influenza cases from a college community. In: *Proceedings of the National Academy of Sciences of the United States of America*, pp 1081–6
6. Delamater PL, Street EJ, Leslie TF et al (2019) Complexity of the basic reproduction number (R0). *Emerg Infect Dis* 25(1):1–4. <https://doi.org/10.3201/eid2501.171901>
7. Dietz K (1993) The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res* 2:23–41
8. Asadi S, Wexler AS, Cappa CD, Barreda S, Bouvier NM, Ristenpart WD (2020) Effect of voicing and articulation manner on aerosol particle emission during human speech. *PLoS One* 15(1):e0227699
9. Abkarian M, Stone HA (2020) Stretching and break-up of saliva filaments during speech: a route for pathogen aerosolization and its potential mitigation. *Phys Rev Fluids* 5:102301(R)
10. Ladefoged P, Johnson K (2011) *A course in phonetics*, 6th edn. Wadsworth/Cengage Learning, Boston
11. Inouye S (2003) SARS transmission: language and droplet production. *Lancet* 362(9378):170
12. Inouye S, Sugihara Y (2015) Measurement of puff strength during speaking: comparison of Japanese with English and Chinese languages. *J Phonetic Soc Jpn* 19(3):43–49
13. Georgiou GP, Kilani A (2020) The use of aspirated consonants during speech may increase the transmission of COVID-19. *Med Hypotheses* 144:109937
14. Worldometers.info (2020). COVID-19 coronavirus pandemic. Dover, Delaware, U.S.A. <https://www.worldometers.info/coronavirus/>
15. Moran S, McCloy D (eds) (2019) PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. Available online at <http://phoible.org>, Accessed on 2020-10-17.
16. Maddieson I (1984) *Patterns of sounds*. Cambridge University Press, Cambridge
17. Michael L, Stark T, Chang W (2012) South American Phonological Inventory Database. <http://linguistics.berkeley.edu/~saphon/en/> accessed 4 February 2017
18. Crothers JH, Lorentz JP, Sherman DA, Vihman MM (1979) *Handbook of phonological data from a sample of the world's languages (A report of the Stanford Phonology Archive)*. Stanford University, Stanford, CA
19. Mielke J (2018) Visualizing phonetic segment frequencies with density-equalizing maps. *J Int Phon Assoc* 48(2):129–154
20. R Core Team (2020) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
21. Peust C (2008) On consonant frequency in Egyptian and other languages. *Lingua Aegyptia* 16:105–134

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.