

Article

Discriminant Analysis of Voice Commands in the Presence of an Unmanned Aerial Vehicle

Marzena Mięsikowska

Faculty of Mechatronics and Mechanical Engineering, Kielce University of Technology, 25-314 Kielce, Poland; marzena@tu.kielce.pl

Abstract: The aim of this study was to perform discriminant analysis of voice commands in the presence of an unmanned aerial vehicle equipped with four rotating propellers, as well as to obtain background sound levels and speech intelligibility. The measurements were taken in laboratory conditions in the absence of the unmanned aerial vehicle and the presence of the unmanned aerial vehicle. Discriminant analysis of speech commands (left, right, up, down, forward, backward, start, and stop) was performed based on mel-frequency cepstral coefficients. Ten male speakers took part in this experiment. The unmanned aerial vehicle hovered at a height of 1.8 m during the recordings at a distance of 2 m from the speaker and 0.3 m above the measuring equipment. Discriminant analysis based on mel-frequency cepstral coefficients showed promising classification of speech commands equal to 76.2% for male speakers. Evaluated speech intelligibility during recordings and obtained sound levels in the presence of the unmanned aerial vehicle during recordings did not exclude verbal communication with the unmanned aerial vehicle for male speakers.

Keywords: speech intelligibility; drone; unmanned aerial vehicle (UAV); sound level



Citation: Mięsikowska, M. Discriminant Analysis of Voice Commands in the Presence of an Unmanned Aerial Vehicle. *Information* **2021**, *12*, 23. <https://doi.org/10.3390/info12010023>

Received: 11 October 2020

Accepted: 6 January 2021

Published: 8 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, the use of unmanned aerial vehicles (UAVs) has increased significantly in various areas of life. Ecologists use UAVs to monitor birds, wildlife, and changes in nature. Recent advances in the use of UAVs for wildlife research are promising and can improve data collection efficiency. UAVs can fly over rough terrain and follow specific trajectories to provide useful data from animals carrying lightweight transmitters. UAVs have great potential for tracking many species [1–5]. The excessive noise of unmanned aerial vehicle (UAV) is the main obstacle to using UAVs for bioacoustics monitoring. Studies to reduce motor and rotor noise of UAVs are very important.

In agriculture, UAVs can be used to monitor plantations and to spray crops. In recent years, UAVs have become essential in precision agriculture. Precision agriculture is the application of geospatial techniques and sensors to identify variations in the field and to deal with these variations using alternative strategies [6–10].

UAVs can be used in emergency medical services for medical transport, i.e., for the delivery of medical products, including blood derivatives and pharmaceuticals, to hospitals, mass accident sites, and coastal ships in times of critical demand. They can provide basic medications against, for example, the venom of snake bites and prevent death. UAVs can be used in disaster relief to rescue victims and provide food, water, and medicine. In addition, they can be used to transport organs in a short time, thus preventing delay due to busy traffic. UAVs can be used for surveillance of disaster sites and biohazard areas, and for research in epidemiology. UAVs have the potential to become reliable platforms for delivery of microbial and laboratory samples, vaccines, emergency medical equipment, and patient transport [11–14].

UAVs can be used in mountain rescue for search and rescue missions. Special platforms are built to meet the environmental requirements for mountainous terrain, such as low temperatures, high altitudes, and strong winds. They use vision and thermal imaging

technologies to search for and rescue missing persons in difficult conditions, in snow-storm, and in the forest, both during the day and at night [15,16]. UAVs can be used in firefighting [17,18]. Equipped with smoke detectors, they can detect hazardous gases in the air, as well as fire. They can be used to monitor vegetation and estimate hydric stress and risk index. UAVs can also be used to detect, confirm, locate, and monitor forest fires. Finally, UAVs are also useful for assessing the effects of a fire, especially on the burnt area. UAVs with special characteristics can assist in extinguishing forest fires (forest firefighting). UAVs are also used in military areas for carrying and transporting loads, monitoring, and patrolling the areas, and tracking an object in hazardous conditions [19].

The unmanned aerial vehicle (UAV) is usually controlled by the operator using an operational panel. Commonly, the operator can control a single UAV at a time. The control system that is based on the operator panel significantly reduces the possibility of controlling many UAVs at a time. Such a control system also does not make the operator mobile. Although the control can be done using gestures, the control by voice commands seems to be more efficient [20]. The operator can move around and simultaneously issue commands to many UAVs. In some applications, such control may turn out to be extremely useful [21,22]. In [21], an attempt was made to classify the voice commands spoken by a female operator and the results of the classification was 100% successful. The microphone was placed 0.3 m from the speaker and 2 m from the UAV while speaking the commands. In [22], the control of many voice-operated unmanned aerial vehicles (UAVs) was undertaken.

Automatic speech recognition (ASR) systems can greatly support the UAV control. They can be used to recognize voice commands and thus obtain voice control. Speech recognition attempts were made for UAV [21–25]. However, the problem of significant reduction in the effectiveness of recognizing speech disturbed by the signal of the unmanned aircraft may arise. Work on improving the efficiency of speech recognition systems in the presence of disturbances, i.e., with noise, is ongoing [26,27]. Searching for coefficients and classifiers that increase the recognition accuracy in noisy conditions is very important and is being considered [28]. Many solutions have been proposed to improve recognition accuracy in noisy environments. The first approach is focused on parameterization methods that are resistant to noise or minimize the effect of noise. Other approaches are based on the adoption of clean models by the noise recognition environment in order to contaminate the models or transform a noisy speech into a clean speech—the noise is removed or reduced by the speech representation or implementation of audio–visual speech recognition methods based on lip detection.

Acoustic methods may allow the assessment of the degree of disturbance of the speech signal by the UAV signal. Among other things, it is useful to assess the sound level emitted by the UAV and to estimate the speech intelligibility in the presence of the UAV, determined based on the frequency characteristics of the UAV signal. Speech intelligibility can be assessed using the International ISO Standard 9921 which specifies the requirements for the performance of speech communication for verbal alert and danger signals, information messages, and speech communication in general [29]. Speech interference level (SIL) is one of the parameters that offer a method to predict and assess speech intelligibility in cases of direct communication [29]. As a result, we can assess the quality of the recorded speech signal and select methods of filtering interferences from the useful speech signal. Acoustic methods allow one to determine the distance at which one can communicate clearly with the UAV. Moreover, the acoustic conditions allow one to determine the noise level accompanying voice commands, and thus to approximate the effectiveness of speech recognition in given conditions. This, in turn, may contribute to the adoption of new algorithms for filtering the useful speech signal from disturbances.

The aim of this study was to perform discriminant analysis of voice commands of male speakers in the presence of the UAV, and to determine acoustic conditions in the presence, as well as the absence, of the UAV. In the case of acoustic conditions, the background sound levels, and speech intelligibility were determined. Therefore, the work considers the

classification of voice commands spoken by male speakers in the presence of the UAV, and the acoustic conditions that prevail in the presence of the UAV.

2. Materials and Methods

2.1. The UAV Used in the Experiment

The UAV used in this experiment is presented in Figure 1.



Figure 1. The unmanned aerial vehicle (UAV) used in the experiment.

The UAV is DJI Mavic Pro. The maximum speed of the UAV is 40 mph (65 kph) in sport mode without wind. The maximum hovering time is 24 min (no wind). The overall flight time is 21 min (normal flight, 15% remaining battery level). The maximum takeoff altitude is 16,404 feet (5000 m). The GPS/GLONASS satellite positioning systems is used. During the experiment, the drone hovered at a height of 1.8 m in the laboratory.

2.2. Acoustic Parameters

The measurements of acoustic parameters were taken as presented in Figure 2.

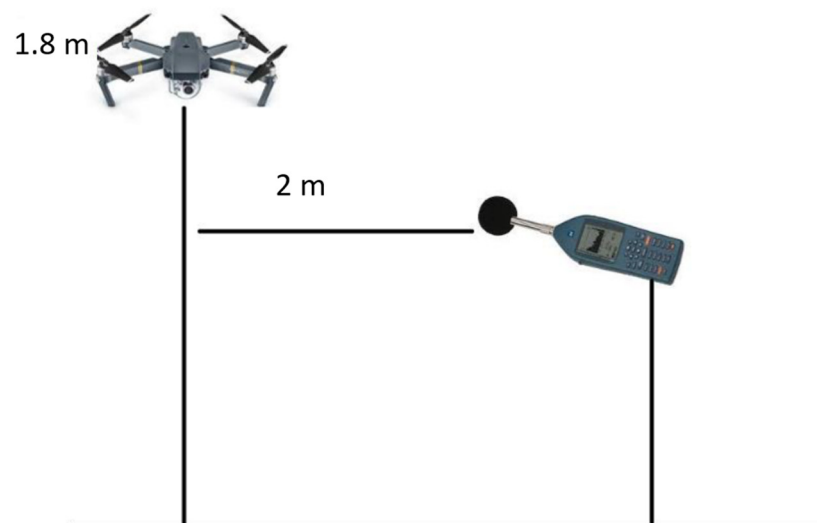


Figure 2. Measurements of acoustic parameters.

Acoustic parameters, such as background sound levels and speech intelligibility, in the presence of the UAV and the absence of the UAV were obtained in this experiment. The

background sound levels were measured using Norsonic 140 sound analyzer in laboratory conditions in the absence of the UAV and the presence of the UAV hovering at a height of 1.8 m at a distance of 2 m from the recording equipment. Speech intelligibility was evaluated based on the speech interference level (SIL) parameter according to the ISO standard 9921 [29]. The SIL parameter is calculated according to Equation (1):

$$\text{SIL} = L_{S,A,L} - L_{\text{SIL}}, \quad (1)$$

where L_{SIL} is the arithmetic mean of the sound-pressure levels in four bands with central frequencies 500 Hz, 1 kHz, 2 kHz, and 4 kHz; and $L_{S,A,L}$ is the level of the speech signal determined by the vocal effort of the speaker [29]. The SIL offers a simple method of predicting or assessing speech intelligibility in cases of direct communication in a noisy environment. It takes into account a simple average of the noise spectrum, the vocal effort of the speaker, and the distance between the speaker and the listener.

2.3. Speakers and Speech Material

Ten male speakers aged 22 to 23 years participated in the study. The speakers' voices were normal; their hearing and speech were not impaired. The native language of the speakers was Polish.

The following speech commands were recorded in laboratory conditions using Olympus LS-11 digital recorder: forward, backward, up, down, left, right, start, and stop. The measurements of speech commands were taken as presented in Figure 3.

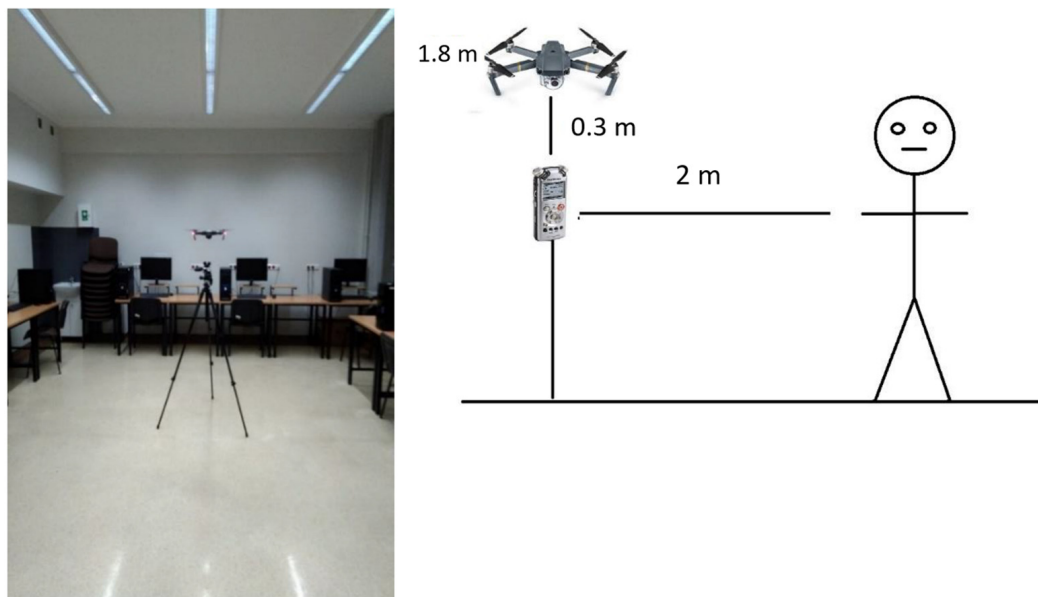


Figure 3. Measurements of speech commands.

The commands were spoken in English. Each command was expressed three times by each speaker. While speaking the commands, the speakers were in a standing position at a distance of 2 m from the recording equipment. The UAV was hovering at a height of 1.8 m above the ground and 0.3 m above the recording equipment.

The LS-11 from Olympus is a linear PCM audio recorder for studio-quality recording in the field. This portable unit can record at rates of up to 24-bit/96 kHz in WMA, MP3, or WAV formats. The recordings of speech commands were obtained at a frequency of 44.1 kHz.

2.4. Time–Frequency Analysis

Time–frequency analysis of recorded speech commands was done for observation purposes. The time–frequency analysis was performed using MATLAB software. The /backward/ command of a randomly selected speaker in a time–frequency representation was presented to indicate the presence of a speech signal against the background noise of the drone.

2.5. Discriminant Function Analysis

Discriminant function analysis was used to investigate the significant differences between the speech commands of male speakers using 12 mel-frequency cepstral coefficients (MFCC) in the presence of the UAV.

2.5.1. Mel-Frequency Cepstral Coefficients

Twelve MFCC were extracted from the speech recordings. The reason for using the MFCC was their application in speech recognition and their very effective classification method in previous experiments using discriminant function analysis when analyzing speech signals recorded in the presence of the UAV for a single female speaker [21].

2.5.2. Discriminant Function Analysis

The speech commands were taken as the grouping variables and the MFCC were taken as the independent variables. The discriminant analysis included the discrimination stage and the classification stage. Discriminant analysis was performed using STATISTICA software [30]. In the discrimination stage, the maximum number of discriminant functions computed was equal to the number of groups minus one. A canonical analysis was performed to determine the successive functions and canonical roots. The standardized coefficients were obtained in each discriminant function. The larger the standardized coefficients, the greater the contribution of the variable to the discrimination between groups. Chi-square tests with successive roots removed were performed. The coefficient of canonical correlation (canonical-R) is a measure of the association between the *i*-canonical discriminant function and the group. The canonical-R ranges between 0 (no association) and 1 (very high association). Wilks' lambda statistic is used to determine the statistical significance of discrimination and ranges between 0 (excellent discrimination) and 1 (no discrimination).

The classification stage proceeded after the variables that discriminate the speech groups were determined. Because there were eight speech groups, eight classification functions were created according to Equation (2):

$$K_i(h) = c_{i0} + w_{i1}mfcc_1 + w_{i2}mfcc_2 + \dots + w_{i12}mfcc_{12}, \quad (2)$$

where *h* is the speech commands considered as a group (backward, down, forward, left, right, start, stop, up); the subscript *i* denotes the respective group; c_{i0} is a constant for the *i*'th group; w_{ij} is the weight for the *j*'th variable in the computation of the classification score for the *i*'th group; and $mfcc_j$ is the observed mel-cepstral value for the respective case. The classification functions can be used to determine to which group each case most likely belongs. A case is classified as belonging to the group for which it has the highest classification score. With such functions, a case is classified under the group for which $K_i(h)$ assumes the highest value. The classification matrix shows the number of cases that were correctly classified and those that were misclassified.

3. Results

In this experiment, the following results of the acoustic parameters, time–frequency analysis, and discriminant analysis were obtained.

3.1. Acoustic Parameters

3.1.1. Background Sound Levels

In Figure 4, the background sound levels recorded with the Norsonic 140 sound analyzer in laboratory conditions in the absence of the UAV (blue line) and the presence of the UAV (red line) are presented.

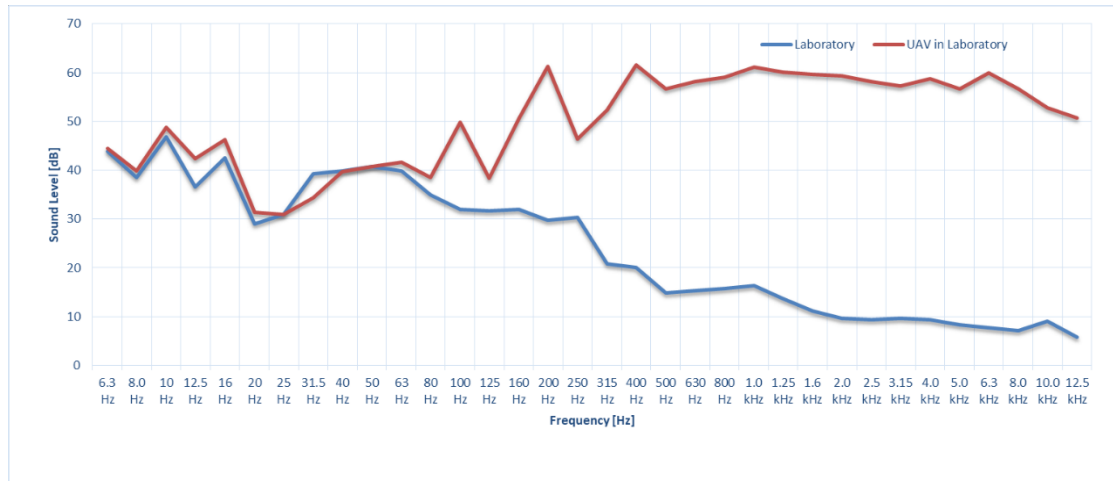


Figure 4. The background sound levels in a laboratory room in the absence of the UAV (blue line) and the presence of the UAV (red line).

The A-weighted sound level was 70.5 dB (A) in the presence of the UAV and 28.4 dB (A) in the absence of the UAV in the laboratory. The characteristic peaks of the UAV appeared at 10 Hz, 16 Hz, 100 Hz, 200 Hz, 400 Hz, 1 kHz, 2 kHz, 4 kHz, and 6.3 kHz.

3.1.2. Speech Intelligibility

The obtained speech intelligibility based on the SIL in laboratory conditions in the presence of the UAV and the absence of the UAV is presented in Table 1.

Table 1. Speech intelligibility in the laboratory in the presence of the UAV and the absence of the UAV.

UAV	L _{Aeq} (dB(A))	L _{S,A,L} (dB)	L _{SIL} (dB)	SIL (dB)	Rating
present	70.5	65.98	58.98	7.0	Poor
absent	28.4	47.98	12.58	35.4	Excellent

When the UAV was present in the laboratory, hovering at a height of 1.8 m, the speech intelligibility was poor. When the UAV was absent, the speech intelligibility was excellent.

3.2. Time–Frequency Analysis

In Figure 5, the time–frequency analysis of the /backward/ speech command, spoken in laboratory conditions in the absence of the UAV and the presence of the UAV, is presented.

In Figure 5a, the /backward/ speech command without any disturbances from the UAV noise is presented. In Figure 5b, the /backward/ speech command with disturbances from the UAV noise is presented. If the speech signal is influenced by the UAV noise, it can still be observed in the recordings. The presence of a speech signal against noise means that an attempt might have been made to classify the voice commands.

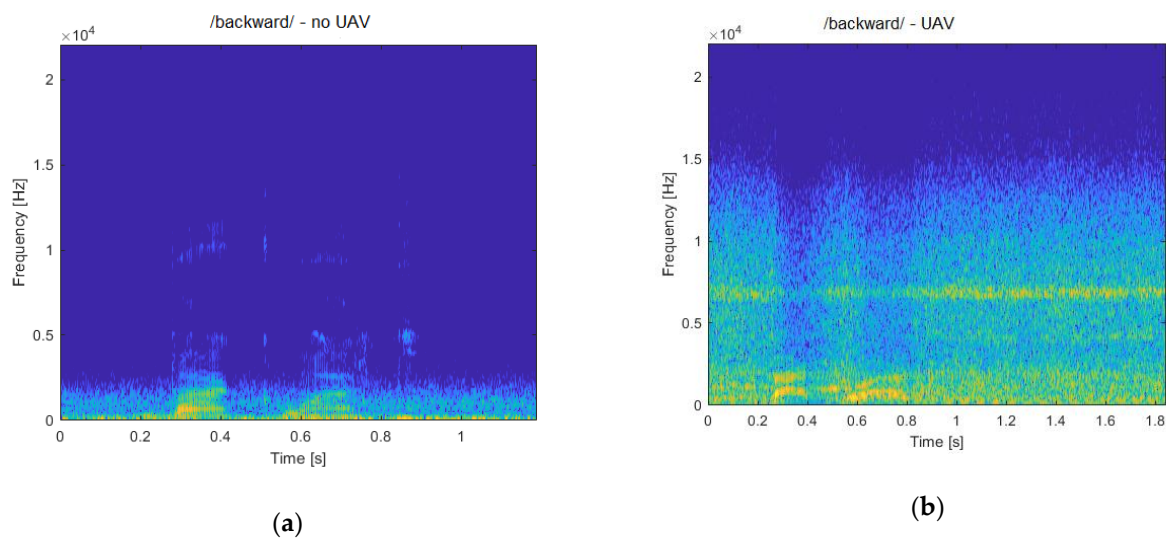


Figure 5. Time–frequency analysis of the /backward/ speech command: (a) in the absence of the UAV, (b) in the presence of the UAV hovering at a height of 1.8 m in laboratory conditions.

3.3. Discriminant Function Analysis

Discriminant function analysis was performed based on 12 MFCC as independent variables and the speech commands as grouping variables, using STATISTICA software. The analysis showed the main significant effects used in the model (Wilks' lambda: 0.0384492; approx. $F(84,1361) = 11.38451$; $p < 0.0000$). Seven discriminant functions (Root1, Root2, Root3, Root4, Root5, Root6, and Root7) were obtained. Chi-square tests with successive roots removed were performed in the canonical stage; they are presented in Table 2.

Table 2. Chi-Square tests with successive roots removed.

Roots Removed	Canonical R	Wilks'-Lambda	Chi-Square	<i>p</i> -Value
0	0.853	0.0384	746.18	0.00000
1	0.807	0.1415	447.78	0.00000
2	0.621	0.4051	206.95	0.00000
3	0.471	0.6593	95.40	0.00000
4	0.343	0.8475	37.89	0.03560
5	0.171	0.9606	9.21	0.81733
6	0.102	0.9895	2.41	0.87868

According to Table 2, chi-square tests with successive roots removed showed the significance of all created discriminant functions used in the model ($R = 0.853$, Wilks' lambda = 0.0384, $p < 0.00000$). The removal of the first discriminant function showed the high canonical value R between the groups and the discriminant functions ($R = 0.807$). The removal of the second, third, fourth, through to the seventh discriminant functions also showed the high canonical value R .

After performing the canonical stage and deriving discriminant functions with 12 MFCC features that discriminate mostly between groups, the study proceeded with the classification stage. The coefficients of the classification functions obtained for the groups are presented in Table 3.

Results of classification using classification functions $K(h)$ for speech commands groups are presented in Table 4.

Table 3. The coefficients of classification functions.

ci	K(backward)	K(down)	K(forward)	K(Left)	K(Right)	K(Stop)	K(Start)	K(up)
wi1	55.38	61.70	52.70	58.95	59.33	52.06	45.64	57.19
wi2	−137.52	−142.67	−137.90	−141.10	−138.87	−134.25	−135.41	−150.02
wi3	333.32	345.43	340.89	332.51	331.93	326.86	335.78	349.35
wi4	−490.28	−492.94	−491.04	−484.72	−493.49	−491.92	−487.04	−492.81
wi5	484.16	474.44	469.49	491.43	491.14	479.79	466.35	459.08
wi6	−360.13	−340.22	−353.48	−359.10	−356.12	−356.51	−364.24	−328.74
wi7	167.30	158.24	156.79	162.39	169.22	170.53	162.14	151.11
wi8	−208.60	−200.27	−214.34	−199.45	−203.70	−203.29	−210.45	−201.63
wi9	262.83	282.21	276.07	262.80	264.91	260.24	275.35	292.98
wi10	−338.17	−340.70	−338.38	−352.07	−342.34	−330.30	−328.55	−343.81
wi11	438.68	445.97	427.19	452.92	440.11	422.68	427.06	445.52
wi12	−415.15	−435.94	−426.25	−409.54	−414.44	−408.60	−420.28	−446.80
cio	−1348.46	−1376.38	−1365.74	−1343.55	−1357.98	−1328.74	−1336.58	−1392.33

For example, the value twenty-nine (29) in Table 4 means that for thirty (30) considered records for each command, twenty-nine were correctly classified under the considered group using the respective classification function $K(h)$. The other values should be read similarly. The value zero (0) means that no record was classified as belonging to the considered group using the function $K(h)$. According to Table 4, the total classification stage was successful (76.2%). The /backward/ command obtained the worst classification rate (43.3%). It was misclassified under the following commands: /forward—four (4) cases, left—four (4) cases, right—four (4) cases, start—three (3) cases, stop—two (2) cases/. Only thirteen (13) cases of the /backward/ command were correctly classified. For each of the remaining commands, the correct classification exceeded more than 20 cases, which resulted in a score equal or above 70%. The /up/ command obtained the best classification rate (96.6%)—twenty-nine (29) cases out of thirty (30) possible cases were correctly classified. Only one (1) case was misclassified as a /right/ command. The values of the columns of the /Total/ row should be interpreted as the number of all cases classified under the given function $K(h)$. A value of 37 means that 37 cases were classified under the function $K(\text{start})$ of all considered cases. The percentage value is the average value of correctly classified cases.

Table 4. The classification matrix.

Group	%	K(backward)	K(down)	K(forward)	K(Left)	K(Right)	K(Start)	K(Stop)	K(up)
backward	43.3	13	0	4	4	4	3	2	0
down	83.3	0	25	2	0	2	0	0	1
forward	70	5	0	21	0	0	1	3	0
left	70	4	0	0	21	5	0	0	0
right	73.3	3	0	1	1	22	3	0	0
start	93.3	1	0	0	0	0	28	1	0
stop	80	3	0	1	0	0	2	24	0
up	96.6	0	0	0	0	1	0	0	29
Total	76.2	29	25	29	26	34	37	30	30

In general, the discriminant analysis showed significant differences between different speech commands influenced by the UAV noise. There were no significant differences between the same commands influenced by the UAV noise.

4. Discussion

Referring to the acquired acoustic parameters, the background sound levels during the hovering of the UAV in the laboratory room were higher than the background sound levels obtained at normal conditions in the laboratory when no UAV was hovering. The A-weighted sound levels were 70.5 dB (A) in the presence of the UAV and 28.4 dB (A) in

the absence of the UAV. The characteristic peaks of the sound level when the UAV was present in the laboratory appeared at 10 Hz, 16 Hz, 100 Hz, 200 Hz, 400 Hz, 1 kHz, 2 kHz, 4 kHz, and 6.3 kHz. Speech intelligibility was excellent in the absence of the UAV and poor in the presence of the UAV in laboratory conditions. The presence of the UAV did not exclude the possibility of verbal communication with it.

Acoustic parameters can play an important role in predicting the degree of classification of voice commands. When the sound level reached 70.5 dB (A) and the intelligibility was poor, the rating level was as high as 76.2% for male speakers. However, an increase in the sound level and a decrease in speech intelligibility may predict the inability to classify voice commands. The acoustic parameters can therefore provide us with quite important information about the environmental conditions. Based on the acoustic parameters, the drone could independently control the position of the microphone for acquiring voice commands in order to successively classify them. However, further research is needed in this area.

Time–frequency analysis showed the frequency band characteristic for the speech commands as well as frequency bands that were characteristic of the UAV. The speech signal was clearly seen against the background noise from the UAV. The clear presence of the speech signal made it possible to take steps to de-noise the useful speech signal from the UAV noise.

Discriminant analysis based on 12 MFCC showed significant differences between the speech command groups. The classification accuracy was 76.2%, which is promising for verbal communication with the UAV and confirms the possibility of such communication for male speakers. Most of the commands were correctly classified and assigned to the appropriate speech group. Only the /backward/ command obtained a classification rate (43.3%) less than 70%. Discriminant analysis and MFCC coped very well with the correct classification of the voice commands.

The results presented in this study showed the possibility of effective communication between the male speakers and the UAV, when the UAV was located directly above the recorder at a distance of 0.3 m and the speaker was at a distance of 2 m from the recorder. The obtained result of 76.2% is satisfactory. The result could be affected by the distance between the UAV and the recorder (interference directly above the recorder) and the fact that the speakers spoke commands in a language other than their native one. The research was performed for male speakers only. A previous study [21] was performed for one female speaker, who gave voice commands in significantly different conditions and position of the recording equipment that resulted in 100% accuracy. In the study [21], the recorder was located at 2 m from the UAV and the speaker was located at 0.3 m from the recorder. The almost direct presence of the speaker at the recorder showed that the classification was 100% effective. The current research for ten male speakers has shown the possibility of effective classification when the noise (UAV) is directly above the recorder and the speaker is 2 m away from the recorder. Therefore, the position of the recorder in relation to the speaker and UAV and the gender of the speakers are completely different compared to the previous research [21].

It is probable that the further the UAV is from the recorder, the lower the chances of the UAV interference reducing the classification efficiency. It is therefore necessary to study other distances of the UAV and the speaker from the recorder and their impact on the efficiency of the classification. The determination of such distances at which the UAV will not interfere with speech can greatly affect the efficiency of classification and processing of such a signal and its subsequent recognition. Research on classification in external conditions for both women and men should be performed.

The discriminant analysis was used here for classification due to the limited number of commands spoken by the male speakers. Speech recognition systems expanded with numerous vocabularies can significantly extend the time of command recognition. Limiting the vocabulary to simple UAV control commands may allow for effective control of the drone, where the time to achieve recognition and accuracy are very important. This solution

also does not require high hardware and computing power. It can be directly applied to the UAV. Voice control can be very important when controlling many drones at the same time [22].

5. Conclusions

The aim of this study was to perform discriminant analysis of voice commands of male speakers in the presence of the UAV, and to determine the acoustic conditions in the presence of the UAV and the absence of the UAV.

For the acoustic conditions, the background sound levels and speech intelligibility were determined. The A-weighted sound level of the UAV hovering at a height of 1.8 m in laboratory conditions was 70.5 dB (A). When the UAV was present in the laboratory, the characteristic peaks appeared at 10 Hz, 16 Hz, 100 Hz, 200 Hz, 400 Hz, 1 kHz, 2 kHz, 4 kHz, and 6.3 kHz. Speech intelligibility was excellent in the absence of the UAV and poor in the presence of the UAV in laboratory conditions during measurements. Evaluating the acoustic parameters, i.e., speech intelligibility and sound level in the drone zone, may be useful for determining the acoustic conditions and the classification level under these conditions. However, obtaining the result of speech intelligibility as /Poor/ did not exclude communication by means of voice commands.

Time–frequency analysis showed characteristic bands for the speech of the male speakers and the characteristic band for the UAV.

Discriminant analysis showed significant differences between the speech command groups. The classification showed 76.2% of accuracy for male speakers. Based on this research, it can be concluded that the position of the UAV and the speaker in relation to the recorder can largely contribute to effective communication with the UAV. If the UAV is directly above the recorder and the speaker is far away, then the UAV can become a significant interference. In this case, this disturbance could have resulted in a classification of voice commands of 76.2% accuracy for men. Further, speaking the voice commands in a language other than the speakers' native one could have contributed to the obtained classification score. Further work in this direction is necessary. Furthermore, other classifiers should be tested.

Future research will be devoted to the study of other distances of the UAV and the speaker from the recorder and their impact on the effectiveness of classification. In addition, external research will be carried out for both women and men. In each of these measurements, acoustic parameters will be tested.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions–privacy. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest

References

1. D'Oleire-Oltmanns, S.; Marzolff, I.; Peter, K.D.; Ries, J.B. Unmanned aerial vehicle (UAV) for monitoring soil erosion in Morocco. *Remote Sens.* **2012**, *4*, 3390–3416. [[CrossRef](#)]
2. Wilson, A.M.; Barr, J.; Zagorski, M. The feasibility of counting songbirds using unmanned aerial vehicles. *Auk* **2017**, *134*, 350–362. [[CrossRef](#)]
3. Paranjape, A.A.; Chung, S.J.; Kim, K.; Shim, D.H. Robotic Herding of a Flock of Birds Using an Unmanned Aerial Vehicle. *IEEE Trans. Robot.* **2018**, *34*, 901–915. [[CrossRef](#)]
4. Tremblay, J.A.; Desrochers, A.; Aubry, Y.; Pace, P.; Bird, D.M. A Low-Cost Technique for Radio-Tracking Wildlife Using a Small Standard Unmanned Aerial Vehicle. *J. Unmanned Veh. Syst.* **2017**, *5*, 102–108. [[CrossRef](#)]
5. Hong, S.J.; Han, Y.; Kim, S.Y.; Lee, A.Y.; Kim, G. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors* **2019**, *19*, 1651. [[CrossRef](#)]

6. Primicerio, J.; Di Gennaro, S.F.; Fiorillo, E.; Genesio, L.; Lugato, E.; Matese, A.; Vaccari, F.P. A flexible unmanned aerial vehicle for precision agriculture. *Precis. Agric.* **2012**, *13*, 517–523. [[CrossRef](#)]
7. Kelly, M.; Vehicle, U.A.; Algorithm, L. Object-based approach for crop Row Characterization in UAV Images for Site-Specific Weed Management. In Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil, 7–9 May 2012.
8. Zhang, C.; Kovacs, J.M. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* **2012**, *13*, 693–712. [[CrossRef](#)]
9. Huang, Y.B.; Thomson, S.J.; Hoffmann, W.C.; Lan, Y.; Fritz, B.K. Development and prospect of unmanned aerial vehicle technologies for agricultural production management. *Int. J. Agric. Biol. Eng.* **2013**, *6*, 1–10.
10. Mogili, U.R.; Deepak, B.B.V.L. Review on Application of Drone Systems in Precision Agriculture. *Procedia Comput. Sci.* **2018**, *133*, 502–509. [[CrossRef](#)]
11. Thiels, C.A.; Aho, J.M.; Zietlow, S.P.; Jenkins, D.H. Use of unmanned aerial vehicles for medical product transport. *Air Med. J.* **2015**, *34*, 104–108. [[CrossRef](#)]
12. Laksham, K. Unmanned aerial vehicle (drones) in public health: A SWOT analysis. *J. Fam. Med. Prim. Care* **2019**, *8*, 342–346. [[CrossRef](#)] [[PubMed](#)]
13. Haidari, L.A.; Brown, S.T.; Ferguson, M.; Bancroft, E.; Spiker, M.; Wilcox, A.; Ambikapathi, R.; Sampath, V.; Connor, D.L.; Lee, B.Y. The economic and operational value of using drones to transport vaccines. *Vaccine* **2016**, *34*, 4062–4067. [[CrossRef](#)] [[PubMed](#)]
14. Rosser, J.C., Jr.; Vignesh, V.; Terwilliger, B.A.; Parker, B.C. Surgical and Medical Applications of Drones: A Comprehensive Review. *JSLS J. Soc. Laparoendosc. Surg.* **2018**, *22*, e2018.00018. [[CrossRef](#)] [[PubMed](#)]
15. Scherer, J.; Yahyanejad, S.; Hayat, S.; Yanmaz, E.; Vukadinovic, V.; Andre, T.; Bettstetter, C.; Rinner, B.; Khan, A.; Hellwagner, H. An autonomous multi-UAV system for search and rescue. In Proceedings of the DroNet 2015-Proceedings of the 2015 Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, Florence, Italy, 18–22 May 2015. [[CrossRef](#)]
16. Silvagni, M.; Tonoli, A.; Zenerino, E.; Chiaberge, M. Multipurpose UAV for search and rescue operations in mountain avalanche events. *Geomat. Nat. Hazards Risk* **2017**, *8*, 18–33. [[CrossRef](#)]
17. Ollero, A.; Martínez-de-Dios, J.R.; Merino, L. Unmanned aerial vehicles as tools for forest-fire fighting. *For. Ecol. Manag.* **2006**, *234*, S263. [[CrossRef](#)]
18. Merino, L.; Caballero, F.; Martínez-De-Dios, J.R.; Maza, I.; Ollero, A. An unmanned aircraft system for automatic forest fire monitoring and measurement. *J. Intell. Robot. Syst. Theory Appl.* **2012**, *65*, 533–548. [[CrossRef](#)]
19. Chamola, V.; Kotesch, P.; Agarwal, A.; Naren; Gupta, N.; Guizani, M. A Comprehensive Review of Unmanned Aerial Vehicle Attacks and Neutralization Techniques. *Ad Hoc Netw.* **2021**, *111*, 102324. [[CrossRef](#)]
20. Draper, M.; Calhoun, G.; Ruff, H.; Williamson, D.; Barry, T. Manual versus speech input for unmanned aerial vehicle control station operations. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting; SAGE Publications: Los Angeles, CA, USA, 2003; Volume 47, No. 1., pp. 109–113.
21. Mięsikowska, M. Speech Intelligibility in the presence of X4 Unmanned Aerial Vehicle. *2018 Signal Process. Algorithms Archit. Arrange. Appl.* **2018**, 310–314. [[CrossRef](#)]
22. Park, J.-S.; Na, H.-J. Front-End of Vehicle-Embedded Speech Recognition for Voice-Driven Multi-UAVs Control. *Appl. Sci.* **2020**, *10*, 6876. [[CrossRef](#)]
23. Bold, S.; Sosorbaram, B.; Batsukh, B.-E.; Ro Lee, S. Autonomous Vision Based Facial and voice Recognition on the Unmanned Aerial Vehicle. *Int. J. Recent Innov. Trends Comput. Commun.* **2016**, *4*, 243–249.
24. Anand, S.S.; Mathiyazaghan, R. Design and Fabrication of Voice Controlled Unmanned Aerial Vehicle. *J. Aeronaut. Aerosp. Eng.* **2016**, *5*, 1–5. [[CrossRef](#)]
25. Prodeus, A.M. Performance measures of noise reduction algorithms in voice control channels of UAVs. In Proceedings of the 2015 IEEE International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), Kiev, Ukraine, 13–15 October 2015; pp. 189–192.
26. Revathi, A.; Jeyalakshmi, C. Robust speech recognition in noisy environment using perceptual features and adaptive filters. In Proceedings of the 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 19–20 October 2017; pp. 692–696.
27. Ondas, S.; Juhar, J.; Pleva, M.; Cizmar, A.; Holcer, R. Service Robot SCORPIO with Robust Speech Interface. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 1–11. [[CrossRef](#)]
28. Zha, Z.L.; Hu, J.; Zhan, Q.R.; Shan, Y.H.; Xie, X.; Wang, J.; Cheng, H.B. Robust speech recognition combining cepstral and articulatory features. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 1401–1405.
29. International Standard ISO 9921:2003. Ergonomics—Assessment of Speech Communication. Publication Date: 2003-10. Available online: <https://www.iso.org/standard/33589.html> (accessed on 7 January 2021).
30. Discriminant Function Analysis—STATISTICA Electronic Documentation. Available online: <https://statisticasoftware.wordpress.com/2012/06/25/discriminant-function-analysis/> (accessed on 7 January 2021).