2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

Vocalics in Human-Drone Interaction

Marc Lieser Ulrich Schwanecke

***Abstract*— As the presence of flying robots continues to grow in both commercial and private sectors, it necessitates an understanding of appropriate methods for nonverbal interaction with humans. While visual cues, such as gestures incorporated into trajectories, are more apparent and thoroughly researched, acoustic cues have remained unexplored, despite their potential to enhance human-drone interaction. Given that additional audiovisual and sensory equipment is not always desired or practicable, and flight noise often masks potential acoustic communication in rotary-wing drones, such as through a loudspeaker, the rotors themselves offer potential for nonverbal communication. In this paper, the consequential sound during the flight of a quadrotor is utilized and modified to carry acoustic information while maintaining the visually perceived flight characteristics. A user study (N=192) demonstrates that acoustically augmenting the trajectories of two aerial gestures with the proposed approach makes them more easily distinguishable. This enhancement contributes to human-drone interaction through onboard means, particularly in situations where the human cannot see or look at the drone.**

I. Introduction

As robots become more integrated into our daily lives, it is critical to ensure effective and safe human-robot interaction (HRI). Failure to meet human expectations can quickly lead to frustration. Poorly implemented robot responses can negatively impact the acceptance of robots, turning otherwise positive encounters into unpleasant experiences.

The field of human-drone interaction (HDI) is dedicated to enhancing the relationship and collaboration between humans and flying robots in various interaction aspects, ultimately improving human-robot encounters in everyday environments. It is essential for robots to comprehend human commands, while humans should intuitively understand a robot's intent. The responsibility lies with robot designers and HRI researchers to craft social robots and provide appropriate interaction methods. Established methods to indicate, e. g., a robot's status through visual cues such as displays, are often applied to newer platforms.

There are, however, robots that are constrained in their appearances and where the adoption of such methods is more challenging. Flying robots face significant limitations in payload capacity, often needing to prioritize equipment for flight-related functions or weight reduction. But despite their mechanical design, they can be retrofitted with features that enhance human acceptance and interaction experience, such as exploiting possible channels of nonverbal communication.

The authors (givenname.surname@hs-rm.de) are with the Computer Vision & Mixed Reality Group of RheinMain University of Applied Sciences, Wiesbaden, Germany.

A. Nonverbal Human-Drone Communication

The field of nonverbal HDI is still emerging. Due to their airborne nature, established concepts from ground-based robots are not directly applicable to flying robots. For example, screens that display facial expressions to convey the drone's emotions have been attached to both physical drones [1] and virtual flying platforms [2]. However, the additional payload significantly shortens the flight time of the physical drones. Focussing on the utilization of the existing possibilities of a drone, kinesics, proxemics, and haptics stand out as the most prominent research topics in the field. Drone *kinesics* researches how emotions or intent can be conveyed through the way the robots move [3], [4], [5], [6], [7] which also includes aerial gestures, e. g., to be utilized for pedestrian guidance [8]. The field of drone *proxemics* examines the extent to which people are comfortable with the proximity of drones and how close they tolerate them to their body [9], [1], [10], [11], [12], [13]. Drone *haptics* is a field of research that studies direct interaction with drones, e. g., the landing of drones on the human body [14], direct control with physical buttons attached to the drone [15], or virtual buttons utilizing the onboard inertial measurement unit and thus dispensing with additional hardware [16].

Research in the area of nonverbal, sound-based communication has increased over the past few years [17], [18]. *Vocalics*, the study of paralanguage, refers to the nonphonemic properties of speech, such as volume, pace, pitch, rate, rhythm, and tone. Paralanguage offers great potential to enhance interaction as it conveys both meaning and emotion through vocal characteristics. Semantic-free utterances (SFUs) in animated films allow robots to communicate affect and intent [19] and inspired researchers to enhance robot communication by artificial movement sound [20]. It also motivated efforts to generalize sound design recommendations [21] based on interviews with the designers of robotic consumer products like the Anki Vector, which uses non-lexical audio for communication [22].

B. Motivation

While nonverbal communication channels such as kinesics and proxemics have been extensively researched, vocalics has not yet found widespread adoption in the field of HDI. One possible reason for this is the rotor sound already emitted as a consequence of the drone's operation, referred to as consequential sound. Being perceived as noise by humans, it negatively affects robot interaction experience [23], [24], [8], [25], [26]. For this reason, among others, the reduction of acoustic signature is the subject of current research, including the development of active noise cancellation solutions [27] and the exploration of new blade geometries such as the toroidal rotor [28]. However, as of now, neither of these technologies has reached the consumer stage.

To incorporate potential forms of acoustic communication, additional sounds must compete with the drone's consequential sound. Adding natural sounds, such as the twittering of birds or rain, through a loudspeaker attached to a drone increased interaction pleasantness but also the perceived loudness [29]. There are, however, groups that can benefit from rotor noise, such as visually impaired people who can be aurally navigated using a quadrotor's rotor sound [30], [31], [32], [33]. For this reason, shaping the consequential sound of quadrotors should be exploited and evaluated for acoustic communication.

Utilizing the noise sources that are already present as a consequence of the operation of rotary-wing drones is a minimalist approach that avoids the additional payload that even a small loudspeaker entails. Supplementing gestures with vocalic cues allows for interactions with the drone without the necessity of keeping it in sight at all times. There are also scenarios where the drone is not intended to be in the human's line of sight, such as filming a jogger from behind.

The possible augmentation of rotor sounds, such as by integrating SFUs similar to those used by animated film characters, has the potential to create a social drone that addresses multiple channels of nonverbal communication, providing a richer interaction experience than what is available today. Realizing plausible vocalic character cues through the rotors of a drone could open the possibility of elevating a companion robot like Vector to an authentic implementation of a flying social robot.

This paper proposes a vocalic approach by augmenting a quadrotor's consequential sound with a higher-frequency oscillation while maintaining the visually perceived flight characteristics. This method can be used to acoustically augment drone flights to better differentiate aerial gestures or, in its simplest form, to attract attention. In this way, a trajectory can convey more information, just as a shaky voice conveys additional auditory cues in a human dialogue. A user study shows that this method can make two similar-sounding trajectories more clearly distinguishable.

C. Structure

The organization of this paper is as follows: Section II describes the drone platform used as well as the positive and negative feedback trajectories including the acoustic extension to better distinguish the negative feedback trajectory. A detailed account of the user study is provided in Section III. Section IV discusses its results and limitations of the proposed method. Finally, Section V concludes this paper and suggests directions for future research.

II. Airborne Gestures

Two very basic human gestures to communicate positive and negative feedback in western civilizations are nodding and shaking one's head. Transferring these motions to a drone results in similar sounding flights, despite their intent being opposite. It can be challenging to identify them based on acoustic cues alone. Especially in scenarios where a person cannot visually track a drone, audibly distinguishing its intent could enhance the interaction experience.

Augmenting the consequential sound of a drone in one of the trajectories promises to improve differentiability. One way to achieve this is to adjust the trajectory controller used. However, in order not to interfere with an established controller, the concrete implementation pursued in this approach is to modify the trajectories that are then fed to the controller.

This section describes the implementation of the basic positive and negative airborne feedback gestures that mimic human head nodding and head shaking. It then addresses the augmentation by the proposed vocalics approach, as well as the drone platform used for the concrete implementation.

A. Trajectories

Quadrotors are underactuated systems, i. e., translational motion in the vertical plane is coupled to rotational motion about their roll and pitch axes. For this reason, the human nod is mimicked by a simple up and down motion, rather than the actual nod being mimicked by a rotation about the pitch axis of the drone, which would introduce horizontal motion. To communicate negative feedback, the drone repeatedly rotates left and right about its yaw axis to mimic head shaking.

The trajectories are based on a simple harmonic motion

$$h(t, f, p) = \sin(2\pi f \cdot t + p)$$

with time t, frequency f, and phase shift p. For the positive feedback trajectory, the scaling of this function models an altitude offset

$$\mathrm{d}z(t) = 0.03 \cdot h(t, 1, \pi)$$

which is applied to the drone's current altitude z. A phase shift of $p = \pi$ is applied to start the nodding motion downward first. The head shaking motion used for the negative feedback trajectory affects the drone's heading angle ψ and has the same characteristics (thus includes the phase shift), scaled to

$$\mathrm{d}\psi(t) = \frac{\pi}{6} \cdot h(t, 1, \pi).$$

The trajectories were generated using a time duration of 4 s. Flying these trajectories using the drone described in Section II-B results in audible sounds that are difficult to distinguish. This similarity is visible in the short-time Fourier transform (STFT) of a microphone recording of the flights performed and is shown in Figs. 1a and 1b. If a quadrotor performs a translational motion along its vertical axis, thrust is equally applied to all four rotors, increasing their frequency (Fig. 1a). When a quadrotor spins about its vertical axis, more thrust is applied to one opposite pair of rotors, while the thrust to the other pair is reduced (Fig. 1b).

Due to the similarity in applied thrust, both flights sound very similar. To better distinguish the negative feedback trajectory from the positive feedback trajectory, the negative feedback trajectory was superimposed by an additional 10 Hz harmonic motion

$$\mathrm{d}\psi_{\mathrm{voc}}(t) = \mathrm{d}\psi(t) + \frac{\pi}{3} \cdot h(t, 10, 0)$$

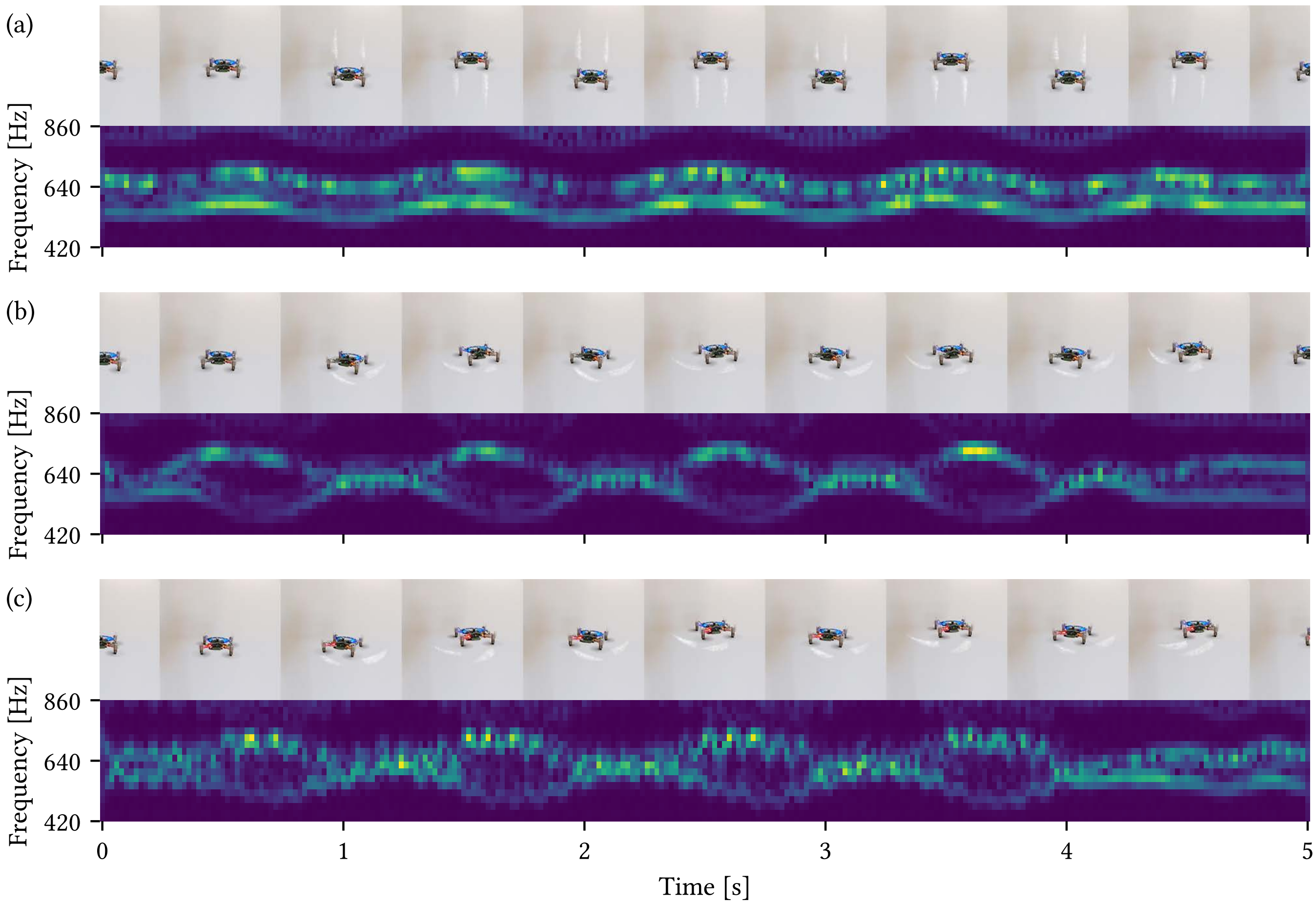

Fig. 1: Stills from trajectories that communicate positive feedback (a), *ordinary* negative feedback (b), and *vocalics* negative feedback (c), and the STFTs of their microphone recordings. The drone's motion is visually emphasized through illustrated traces. The measured frequencies are double the rotor frequencies, since two-blade rotors are used.

to obtain the *vocalics* negative feedback trajectory. The STFT of the sound recorded during its flight is shown in Fig. 1c.

The relevant coordinates of both the *ordinary* negative feedback trajectory and the *vocalics* negative feedback trajectory are shown in Fig. 2b and Fig. 2c, respectively. The displayed heading angle obtained by the pose estimation system during flight does not reflect the oscillation applied to the reference trajectory. This underscores the fact that despite the superimposition of the clearly audible higher-frequency signal on the *vocalics* negative feedback trajectory, the visual impression of the flight is not significantly altered compared to the *ordinary* negative feedback trajectory.

B. Drone Platform

For this experiment, the slightly modified Bitcraze Crazyflie 2.1 platform shown in Fig. 3 is used. It is controlled from a regular laptop running an implementation of a commonly utilized control algorithm [34] using a Crazyradio PA 2.4 GHz USB dongle. The quadrotor is turned upside down to be extended by an active infrared LED tracking marker, which is tracked by a monocular pose estimation system [35]. With the specific implementation it provides poses with six degrees of freedom at a rate of 100 Hz. In the specific setup for this experiment, a single Ximea MQ013MG-ON USB3 high-speed vision camera with a resolution of 1280×1024 pixels is used in combination with a Fujinon DF6HA-1B 1:1.2/6 mm lens and a MidOpt FIL BN850/27 bandpass filter matching the LEDs' range of wavelengths. To reduce noise and jitter, a first-order filter with time constant $\tau = 0.05\,\mathrm{s}$ is used to smooth poses over time. This is part of a testbed developed over recent years, enabling high-speed drone maneuvers with high accuracy. However, when flying basic trajectories, such as those employed in this study, commercially available pose estimation solutions tailored for the Crazyflie can be utilized.

III. User Study

To evaluate whether the vocalic extension of gestures can add value in HDI by helping to audibly differentiate similar sounding trajectories, the flights of the positive feedback trajectory, the *ordinary* negative feedback trajectory, and the *vocalics* negative feedback trajectory were recorded. Video and sound recordings were presented separately within an online

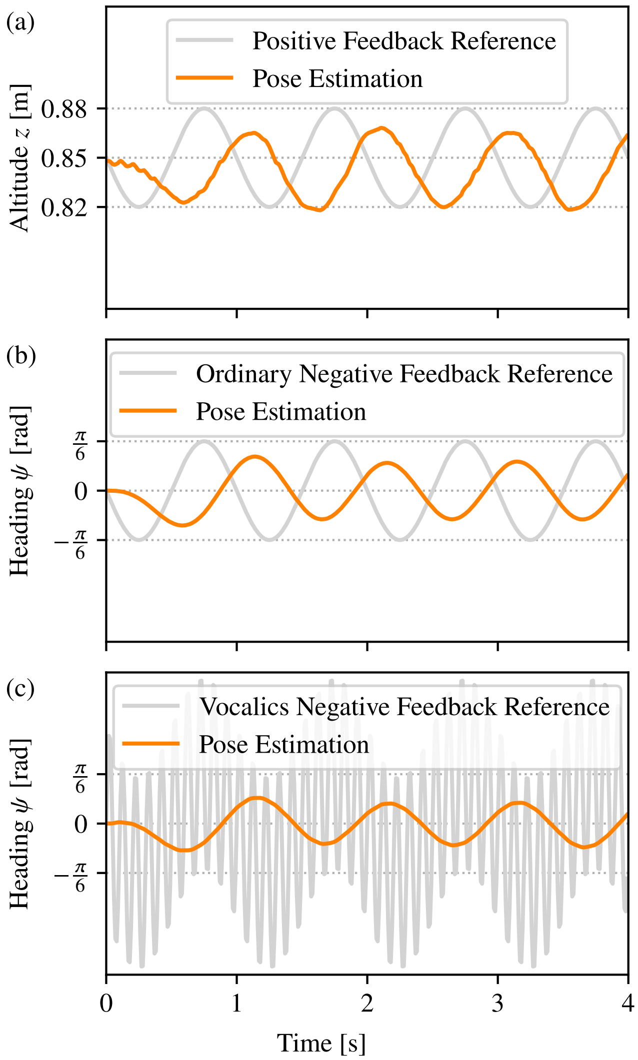

Fig. 2: Relevant reference and tracked parameters of the three trajectories: The altitude z for the positive feedback trajectory (a), and the drone's heading angle ψ for both the *ordinary* negative feedback trajectory (b) and the *vocalics* negative feedback trajectory (c). The background represents the corresponding reference parameter, while the parameter tracked by the pose estimation system is shown in orange.

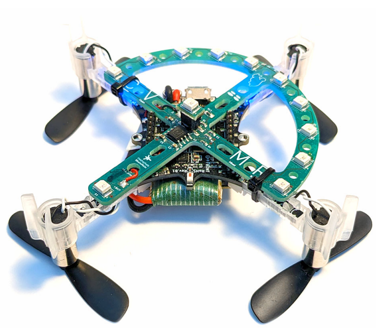

Fig. 3: The Bitcraze Crazyflie 2.1 development drone platform used for this user study, with its rotors mounted upside down to avoid obscuring the LEDs of the tracking marker.

survey to two groups of voluntary participants. The study was implemented using the open source platform oTree [36] and is described in the following.

A. Multimedia Recordings and Synchronization

The videos were recorded using a Google Pixel 5 with default camera settings. The audio was captured with a Shure SM57 instrument microphone equipped with a noise protection filter and a Focusrite Scarlett 2i2 studio-quality interface. Audio and video of the individual airborne gestures were synchronized with a signal tone that was played mid-air just before the start of the gesture. For later evaluation, the software controlling the drone logged important variables at 100 Hz, such as control input, position, attitude, battery voltage, and control method. The latter changes when switching from hover control to trajectory control used for the airborne gestures. This switch is used to synchronize the log data with the audio recording and the signal tone mentioned before.

B. Hypotheses

The alternative hypothesis is that *the group whose negative feedback gestures were acoustically augmented by vocalics can better link the sounds to the videos*, implying higher audible information content in the drone's *vocalics* negative feedback trajectory, as it is visually almost indistinguishable from the *ordinary* negative feedback trajectory. To reject the null hypothesis that *the additional information will not increase the chance of linking the sound correctly to the gestures*, the common significance level of 5 % is chosen.

C. Study Design

The participants were shown a brief written introduction to the user study and a consent form. It was explained that in this study they would be shown two short videos, each showing a drone performing a specific movement. Afterwards, two audio files would be played for them. Furthermore it was stated, that it would be a prerequisite for this study that their device has an audio output and that they should make sure, that the media sound was enabled. After agreeing to the consent form, age, gender, and occupation were asked before the actual experiment started.

The user study was divided into two groups: the *ordinary* trajectory group and the *vocalics* trajectory group. Both groups would see two videos and then listen to two audio files of the same drone flights: the positive feedback trajectory displayed in Fig. 1a and either the *ordinary* negative feedback trajectory shown in Fig. 1b or the *vocalics* negative feedback trajectory displayed in Fig. 1c. The videos presented to participants in the first part of the study (from which the audio for the second part of the study was extracted) are available as supplementary material and on YouTube[1]. To ensure impartiality, participants in both groups were further divided into four subgroups of 24 participants each, so that the media could be presented in the pseudorandom sequence shown in Table I.

[1] `https://www.youtube.com/playlist?list=PLPhCA5Y9lesmAPD5TRXISybvvMxNVeOjA`

TABLE I: Media orders of subgroups where media index 1 is the positive feedback trajectory and media index 2 is the negative feedback trajectory.

Subgroup	Survey Media Order			
	1	2	3	4
	Video	Video	Sound	Sound
1	1	2	1	2
2	1	2	2	1
3	2	1	1	2
4	2	1	2	1

The participants could watch the videos multiple times or listen to the sounds multiple times, but they could not navigate back to the previous media item. The buttons to move to the next page were disabled until the media playback finished. This ensured that the participants played the media at all, and to the end. The data of the participants who have not reached the last page of the study have been deleted.

After each of the two audio files, the participants should link the sound they just heard to one of the two previously shown videos. They had the choice between *Video 1*, *Video 2*, and *Not sure*. Both sounds had to be correctly assigned to the corresponding videos to be considered *correctly linked*.

D. Participants

The user study ($N = 192$) was conducted with 99 males (51.6 %), 90 females (46.9 %), three diverse or undisclosed (1.5 %) participants aged 19 to 69 years (mean $\bar{x} = 36.0$, standard deviation $\sigma = 7.8$). They were asked to voluntarily take part in the study via social media.

E. Results

Out of the 96 participants in the *ordinary* group, 44 (45.83 %) could link both sounds correctly to the gestures whereas for the *vocalics* group, 64 of the 96 participants (66.67 %) were able to do so. The collected data are tabulated as the 2×2 contingency Table II.

TABLE II: Contingency table presenting the results of the conducted user study.

Correctly linked	Ordinary (Group A)	Vocalics (Group B)	Combined response
Yes	44	64	108
No	52	32	84
Total	96	96	192

To evaluate the contingency table, `SciPy`'s [37] implementation of the Barnard's exact test [38] is used. Under the null hypothesis that the additional information introduced by the vocalics will not increase the chance of correct linkage to the gestures, the likelihood of obtaining test outcomes at least as extreme as the observed data is approximately $p = 0.0019$. As the p-value is below the chosen significance level of 5 %, there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.

IV. Discussion

In this section, the outcomes and methodologies of the experiment are analyzed, focusing on how the superimposition of frequencies on the consequential drone sound impacts participants' ability to associate these sounds with aerial gestures. The advantages and limitations of conducting online studies for HDI research are discussed, along with specific technical limitations encountered during the experiment and the constraints of drone vocalics in different environments.

A. Results

The results of the experiment suggest that the superimposition of a frequency to the flight sound in the *vocalics* group had a positive impact on participants' ability to connect the flight sounds to the aerial gestures of the drone. The results indicate that sonically augmenting the trajectory in the *vocalics* group made it easier for participants to associate and differentiate between the flight sounds and gestures, resulting in a higher success rate in linking them compared to the *ordinary* group. This underscores the notion that extending the consequential sound of drones has the potential to convey information, thus opening up a previously unexplored channel of nonverbal communication in HDI.

B. Methodology

Online studies facilitate reaching a larger number of individuals across a broad demographic spectrum. More importantly, impacts typically associated with drone proxemics, such as increased mental stress, arise from the physical presence of the drone. Through an online study, it was ensured that the results were free from these influences. Participants were able to fully focus on the visual and acoustic content of the study in an environment of their own choice.

The online study was conducted independently by the participants without supervision. Control over the playback of multimedia content was thus limited. As described in Section III-C, measures were taken to ensure that the media presented was played to the end. However, browsers do not have access to system settings such as volume. It is assumed that the participants, all of whom volunteered, actually turned on the sound as they were asked to do at the beginning of the survey. There was also no control, if the participants used their device's speaker, headphones, or external speakers. It is anticipated that they are familiar with their devices and consume media on them, having acquired them through social media. If participants did not change their surroundings between individual aerial gestures, there are no concerns in this regard for this study, as the intention was to compare sounds with each other. It is only important that they hear them through the same device and in the same environment. There were two participants (standard score > 2) who took more than the average 2.7 min for the four pages containing media (namely 44 min and 264 min). It was doubtful that they had consumed all media in the same environment. Excluding these two participants from the statistical analysis further decreases the p-value to $p = 0.0017$.

C. Drone Vocalics Limitations

The incorporation of drone rotor sounds in HRI is naturally confined to distances within audible range. Loud environments, where ambient noise drowns out the drone sound are also unsuitable for drone vocalics as an isolated communication channel. On larger drone platforms with slower rotor speeds, it remains to be investigated how the proposed method can be implemented without affecting flight characteristics. The impact of superimposing the trajectory with a higher-frequency signal on the quadrotor battery also remains uncertain and requires further evaluation.

D. Technical Limitations

The yaw motion of the drone induced a noticeable change in altitude in both the *ordinary* and *vocalics* negative feedback trajectories (Figs. 1b and 1c). Rotation about the yaw axis is achieved by torque imbalance. In a quadrotor, opposing rotors spin in the same direction: one pair rotates clockwise and the other pair rotates counterclockwise. To initiate yaw motion, the speed of rotors spinning in one direction is increased, while the speed of rotors spinning in the opposite direction is decreased. This results in a net torque, causing angular acceleration about the yaw axis in the desired direction. As rotors spin up faster than they spin down, the sum of the forces generated by the individual rotors during the yaw motion can momentarily exceed the thrust necessary to maintain the current altitude. In addition, the flight controller can only reduce the rotor speed to a minimum idle speed, which contributes to this effect.

There is a latency of the flown trajectories to their reference trajectories visible in Fig. 2. This does not affect the study or its results, but the latency should be further reduced by optimizing the controller parameters as well as the pose estimation system parameters.

The total rotation angle of the yaw motion was set to 60° in the reference negative feedback trajectory. As the controller used does not take the mathematical model parameters of the specific drone and the feasibility of trajectories into account, the *ordinary* negative feedback trajectory only reached a full head shaking range of 42°. The overlaid signal in the *vocalics* negative feedback trajectory further reduced this range to 32°. This could be compensated by tuning the controller or by aiming for a larger heading angle.

V. Conclusion & Future Work

In this paper, a user study was conducted to evaluate the potential of adding audible information to drone trajectories. Positive and negative feedback trajectories mimicking human gestures were generated and tracked by a quadrotor resulting in visually clearly differentiable flights. However, in situations where keeping the drone in sight is not possible or for visually impaired people, distinguishing airborne gestures acoustically becomes challenging. Superimposing a higher-frequency signal to the negative feedback trajectory did not affect the visual impression of the flight, but could add enough acoustic information to be better differentiable from the positive feedback trajectory. Next to the results and methodology of the user study, the technical limitations have been discussed. Enhancements of the used infrastructure include optimizing pose estimation and controller parameters to further reduce lag and stabilize the drone's altitude during yaw motions.

The presented method represents an initial step toward integrating vocalics into HDI by adding auditory cues utilizing onboard means. Building on this, new ways to further improve HDI can be explored, e. g., by developing new interface designs that include both visual and auditory cues. Especially for individuals with visual impairments, this could make drone technology more accessible.

In future scenarios, where drones and their sounds may have become more commonplace, vocalics could be an additional channel to attract attention or to communicate with pedestrians. This study served as a proof of concept; future studies will focus on exploring a wider variety of tones, communicating more meaningful messages and emotions, and evaluating them qualitatively. The proposed method can be expanded by superimposing different frequencies, encoding sequences of varying signal durations, such as Morse code, or mimicking bio-inspired sounds like the pulsed vibration used by honey bees for communication. Since the proposed approach changes the trajectory instead of the controller parameters, more complex signals can easily be superimposed on the drone's trajectory. Such complex signals could, for example, be SFUs mapped from selected nonphonemic properties of speech, helping to create an authentic implementation of a flying companion robot.

HRI thrives by incorporating as many nonverbal channels as possible. Providing the community with the drone vocalics concept offers a tool that can be used to further enhance the interaction experience. The use and intentional modification of rotor sounds, beyond the sound itself (e. g., for navigation purposes), represents an entirely novel approach to nonverbal communication from drones to humans. In comparison to synthetic acoustic feedback from drones, such as buzzers, the proposed vocalics approach holds potential for more naturally appearing drone feedback, as humans may interpret vocalic communication as a plausible character cue. By adding vocalics to common nonverbal communication channels, such as kinesics, a social drone character could be created that resembles well-known animated film characters and allows for even richer interactions.

References

[1] A. Yeh, P. Ratsamee, K. Kiyokawa, Y. Uranishi, T. Mashita, H. Takemura, M. Fjeld, and M. Obaid, "Exploring proxemics for human-drone interaction," in *Proceedings of the 5th International Conference on Human Agent Interaction*, ser. HAI '17, 2017, pp. 81–88.

[2] V. Herdel, A. Kuzminykh, A. Hildebrandt, and J. R. Cauchard, "Drone in love: Emotional perception of facial expressions on flying robots," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021.

[3] J. R. Cauchard, K. Y. Zhai, M. Spadafora, and J. A. Landay, "Emotion encoding in human-drone interaction," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 3 2016, pp. 263–270.

[4] K. D. Karjalainen, A. E. S. Romell, P. Ratsamee, A. E. Yantac, M. Fjeld, and M. Obaid, "Social drone companion for the home environment: A user-centric exploration," in *Proceedings of the 5th International*

Conference on Human Agent Interaction, ser. HAI '17, 2017, pp. 89–96.

[5] H. Deng, J. Li, A. Sayegh, S. Birolini, and S. Andreani, “Twinkle: A flying lighting companion for urban safety,” in *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '18, 2018, pp. 567–573.

[6] M. Lieser, U. Schwanecke, and J. Berdux, “Eave: Emotional aerial vehicle evaluator,” in *INFORMATIK 2020*, R. H. Reussner, A. Koziolek, and R. Heinrich, Eds. Bonn: Gesellschaft für Informatik, 2021, pp. 899–906.

[7] A. Bevins and B. A. Duncan, “Aerial flight paths for communication: How participants perceive and intend to respond to drone movements,” in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 16–23.

[8] A. Colley, L. Virtanen, P. Knierim, and J. Häkkilä, “Investigating drone motion as pedestrian guidance,” in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 143–150.

[9] B. A. Duncan and R. R. Murphy, “Comfortable approach distance with small unmanned aerial vehicles,” in *2013 RO-MAN*, 2013, pp. 786–792.

[10] U. Acharya, A. Bevins, and B. A. Duncan, “Investigation of human-robot comfort with a small unmanned aerial vehicle compared to a ground robot,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 9 2017, pp. 2758–2765.

[11] A. Wojciechowska, J. Frey, S. Sass, R. Shafir, and J. R. Cauchard, “Collocated human-drone interaction: Methodology and approach strategy,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 3 2019, pp. 172–181.

[12] M. Lieser, U. Schwanecke, and J. Berdux, “Evaluating distances in tactile human-drone interaction,” in *30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, 2021, pp. 1275–1282.

[13] S. Kunde, N. Simms, G. Uriarte, and B. Duncan, “Let's run an online proxemics study! but, how do results compare to in-person?” in *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I.* Berlin, Heidelberg: Springer-Verlag, 2023, p. 24–37.

[14] J. Auda, M. Weigel, J. R. Cauchard, and S. Schneegass, “Understanding drone landing on the human body,” in *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, ser. MobileHCI '21. New York, NY, USA: Association for Computing Machinery, 2021.

[15] S. Rajappa, H. Bülthoff, and P. Stegagno, “Design and implementation of a novel architecture for physical human-uav interaction,” *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 800–819, 2017.

[16] M. Lieser, U. Schwanecke, and J. Berdux, “Tactile human-quadrotor interaction: Metrodrone,” in *15th International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '21. Association for Computing Machinery, 2021.

[17] B. J. Zhang and N. T. Fitter, “Nonverbal sound in human-robot interaction: A systematic review,” *J. Hum.-Robot Interact.*, vol. 12, no. 4, 12 2023.

[18] H. Pelikan, F. A. Robinson, L. Keevallik, M. Velonaki, M. Broth, and O. Bown, “Sound in human-robot interaction,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 706–708.

[19] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, “Review of semantic-free utterances in social human–robot interaction,” *International Journal of Human–Computer Interaction*, vol. 32, no. 1, pp. 63–85, 2016.

[20] F. A. Robinson, M. Velonaki, and O. Bown, “Smooth operator: Tuning robot perception through artificial movement sound,” in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 53–62.

[21] F. Robinson, O. Bown, and M. Velonaki, “Designing sound for social robots: Candidate design principles,” *International Journal of Social Robotics 14*, vol. 14, p. 1507–1525, 06 2022.

[22] H. R. M. Pelikan, M. Broth, and L. Keevallik, “”are you sad, cozmo?”: How humans make sense of a home robot's emotion displays,” in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 461–470.

[23] J. R. Cauchard, J. L. E, K. Y. Zhai, and J. A. Landay, “Drone & me: An exploration into natural human-drone interaction,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '15. ACM, 2015, pp. 361–365.

[24] B. Jones, K. Dillman, R. Tang, A. Tang, E. Sharlin, L. Oehlberg, C. Neustaedter, and S. Bateman, “Elevating communication, collaboration, and shared experiences in mobile video through drones,” in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, ser. DIS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1123–1135.

[25] V. Chang, P. Chundury, and M. Chetty, “Spiders in the sky: User perceptions of drones, privacy, and security,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: ACM, 2017, pp. 6765–6776.

[26] P. Knierim, T. Kosch, A. Achberger, and M. Funk, “Flyables: Exploring 3d interaction spaces for levitating tangibles,” in *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 329–336.

[27] M. D. Narine, M. D. Narine, A. G. Bourgeois, and A. A. Y. Mussa, “Active noise cancellation of drone propeller noise through waveform approximation and pitch-shifting,” Master's Thesis, Georgia State University, Atlanta, Georgia, 5 2020.

[28] T. Sebastian and C. Strem, “Toroidal propeller,” U.S. Patent US10836466B2, 2020.

[29] Z. Wang, Z. Hu, B. Rohles, S. Ljungblad, V. Koenig, and M. Fjeld, “The effects of natural sounds and proxemic distances on the perception of a noisy domestic flying robot,” *J. Hum.-Robot Interact.*, vol. 12, no. 4, 12 2023.

[30] M. Avila Soto, M. Funk, and N. Henze, “Dronenavigator: Using drones for navigating visually impaired persons,” in *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ser. ASSETS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 327–328.

[31] M. Al Zayer, S. Tregillus, J. Bhandari, D. Feil-Seifer, and E. Folmer, “Exploring the use of a drone to guide blind runners,” in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 263–264.

[32] M. Avila Soto, M. Funk, M. Hoppe, R. Boldt, K. Wolf, and N. Henze, “Dronenavigator: Using leashed and free-floating quadcopters to navigate visually impaired travelers,” in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 300–304.

[33] M. Avila Soto and M. Funk, “Look, a guidance drone! assessing the social acceptability of companion drones for blind travelers in public spaces,” in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 417–419.

[34] D. Mellinger and V. Kumar, “Minimum snap trajectory generation and control for quadrotors,” in *2011 IEEE International Conference on Robotics and Automation*, ser. ICRA '11, 5 2011, pp. 2520–2525.

[35] H. Tjaden, “Robust monocular pose estimation of rigid 3d objects in real-time,” PhD Thesis, Johannes Gutenberg University Mainz, Germany, 2019, https://nbn-resolving.org/urn:nbn:de:hebis:77-diss-1000025478. Retrieved: 2021-01-14.

[36] D. L. Chen, M. Schonger, and C. Wickens, “otree—an open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, vol. 9, pp. 88–97, 2016.

[37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.

[38] G. A. BARNARD, “SIGNIFICANCE TESTS FOR 2×2 TABLES,” *Biometrika*, vol. 34, no. 1-2, pp. 123–138, 1 1947.