Feedup, Feedback, and Feedforward in Curve Mid-Air 3D Gestures

Nicolas Burny[1],
Suzanne Kieffer[2],
Nathan Magrofuoco[1],
Jorge Perez-Medina[1],
Paolo Roselli[3],
Jean Vanderdonckt[1],
Alain Vas[1]
Université catholique de Louvain
[1]Louvain Research Institute in Management and Organizations (LouRIM) - Place des Doyens, 1
[2] Institut Langage et Communication (IL&C),
Ruelle de la Lanterne magique 14
[3]Institut de recherche en mathématique et physique (IRMP) - Chemin du Cyclotron, 2
B-1348 Louvain-la-Neuve, Belgium
{nicolas.burny, suzanne.kieffer, nathan.magrofuoco, jorge.perezmedina, paolo.roselli, jean.vanderdonckt, alain.vas}@uclouvain.be

Abstract
Issuing a mid-air gesture in a three-dimensional space intrinsically suffers for the lack of explicit direct representation of the gesture with which guidance and feedback can be offered. To address this challenge, we decompose the feedback problem into three components: feedup to constantly represent the goal of the gestural task, feedback to respond to what the end user already did related to the initial goal, and feedforward to modify the representation towards the ultimate goal before terminating the gesture production. We exemplify these three components with case studies representing three levels of complexity of Curve Mid-Air 3D Gestures produced in three environments.

Author Keywords
Feedback, feedforward, feedup, gesture interaction, gesture recognition, 3D gesture.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

Introduction
We are interested in the use of curve mid-air 3D gesture in gesture interaction, which are single-path, single stroke continuous gestures produced in mid-air by tracking a particular point (e.g., a finger, the tip of a pointer of an object) with contact sensor (e.g., a 3D trackpad) or without (e.g., by camera-based computer

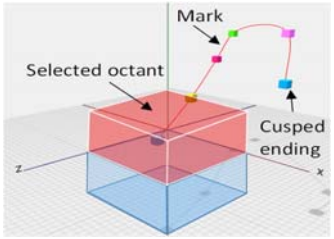

Figure 1. A simple 3D gesture made up of a mark and a cusped ending for selecting a cube octant in BouquetMenu.

Figure 2. A medium 3D gesture captured by webcam object tracking.

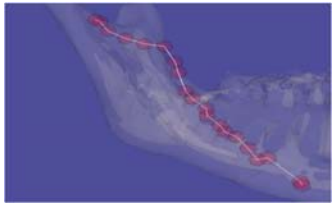

Figure 3. A complex 3D gesture representing a 3D curve for mandible operation in maxilla-facial surgery.

vision). Three levels of gesture complexity are distinguished depending on the curve complexity:

1. *Simple gestures*: gestures having only straight lines in one or two directions. For example, BouquetMenu (Fig. 1) consists of a marking menu offering flicks and marks from an origin towards the directions of the eights octants of a cube. As such, it generalizes into three dimensions the FlowerMenu [1], a marking menu which offers straight marks as gestures into the eight cardinal directions of a compass and which can optionally be terminated by bended, cusped, and pigtail endings.
2. *Medium gestures*: gestures with a given shape representing a defined meaning. For example, a helical spiral represents an air wheel gesture for dimming up or down the light of a room (Fig. 2).
3. *Complex gestures*: gestures which do not have any recognizable shapes or patterns, but represent a curve to be acquired and reproduced. For example, a curve determines how to operat a patient in maxillo-facial surgery after an accident. The system has determined the optimal curve to cut in the jaw of a patient and the surgeon has to exercise to reproduce this gesture until a precision threshold is obtained to minimize invasive surgery (Fig. 3). When reached, the operation can take place either in a collocated room or in teleoperation.

While 3D gesture recognition is now affordable and efficient enough to be incorporated in real-time applications [14], it suffers from an intrinsic problem: there is no 3D representation that would be subject to a direct manipulation for satisfying usual usability heuristics. One such heuristic is “the UI should provide continuous guidance” [2] by answering three questions: where am I?, where do I come from?, and where could I go?. A second equally important heuristic is “Provide immediate feedback” [2,3]: as soon as a gesture has been initiated, the user interface should provide the end user with some visible result indicating that the gesture has been properly recorded so far and that it will undertake the required actions associated to its recognition.

This paper contributes to addressing this problem by decomposing the feedback problem into three components, i.e. feedup, feedback, and feedforward, and exemplifies how to apply them on curve 3D mid-air gestures.

Related work

Regarding a 2D gesture, e.g., a finger on a touch screen (direct representation because the input device equals the output device) or a pen on a remote tablet (indirect representation), it is always possible to graphically represent the gesture path by plotting a trace and to rely on this representation to satisfy the two aforementioned usability heuristics [4] while recognizing such gestures [1,3]. Teaching the motion of a 2D gesture by providing immediate feedback from the recognizer has been revealed effective and efficient [7]. For instance, Octopus [3] graphically suggests completed gestures while the end user is issuing any arbitrary gesture so as to facilitate guidance and feedback and the help end users to learn and remember gestures.

Regarding a 3D gesture, a 3D plotting of the curve captured in mid-air is still possible but this representation will always remain *indirect*: the end user cannot benefit from direct manipulation of this curve since it is often represented as a 3D scene in 2D or a projection on a 2D surface. Because of this lack of guidance and feedback, Octopodus3D [5] generalizes the Octopus [3] idea from 2D to 3D by representing pipes in which 3D

Figure 4: 2D gesture guidance in Octopus [3].

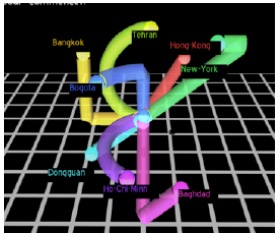

Figure 5: 3D gesture guidance in OctoPodus 3D [5].

gestures can be also terminated. This lack of guidance and feedback can be compensated in various ways, such as haptic feedback for mid-air gestures by ultrasound [9]. For example, AIREAL [11] is a haptic device delivering tactile sensations in mid-air by generating air vortex, thus no longer requiring the user to wear any physical device. Providing such a tactile feedback to mid-air gestures can significantly improve the corresponding gesture interaction [11]. A touchable 3D representation can also be produced in mid-air by combining mixed-reality smart glasses with a haptic device consisting of an array of 16x16 ultrasonic transducers [8]. In conclusion, in order to provide a direct 3D representation of a mid-air gesture that would support guidance and feedback, two options are considered:

1. Keep the constraint of 2D representation of a 3D gesture: in this case, this representation should be augmented with other modalities and/or interaction techniques, such as thermal and vibrotactile feedback in Multi-emoji [16].
2. Relax the constraint of a 2D projection and come up with a direct 3D representation, which can also be augmented with the modalities and interaction techniques that have been positively identified for 2D representations.

The Three Components of Feedback

The power of feedback has been analyzed [6] by decomposing it into three components:

1. *Feedup*: should clarify the goal of a task and establish a clear purpose. When people understand the ultimate goal, they are more likely to focus on the task at hand.
2. *Feedback*: should directly relate to the goal by expressing to what extent any progress has been made with respect to it.
3. *Feedforward*: should modify the goal representation and instructions to plan any future action.

Feedback is certainly the most commonly recognized and extensively researched component in general and for gesture interaction [1,2,5,13], whereas feedforward [15] has been recently introduced successfully [3]. Based on this definition, we revisit these three components by referring to gesture feedback as follows.

Feedup

Feedup should convey the goal of gesture acquisition, based on the three levels of complexity by presenting any relevant information towards this goal. While the end user is issuing any gesture, it could convey information regarding its position in space, its velocity (the first derivative of the position with respect to time), its acceleration (the second derivate of the position with respect to time), its jerk (the third derivate of the position with respect to time),… Higher order derivatives or other features, e.g., tangential acceleration, angular momentum, could also be considered but become harder to geometrically interpret together. For a simple gesture, only the direction matters and the feedup could therefore be limited to indicting this direction. But for a complex gesture, such as in maxilla-facial surgery, velocity matters as well as precision. Velocity can be considered as a vector with its magnitude and direction. Velocity can be computed either instantaneously, which is relevant for feedup, but also on the average, which will be more relevant for feedback.

Figure 6: No feedforward.

Figure 7: Virtual feedforward.

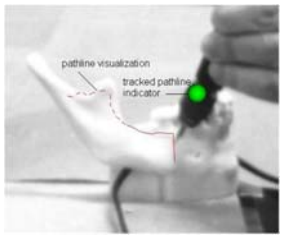

Figure 8: Partially augmented feedforward.

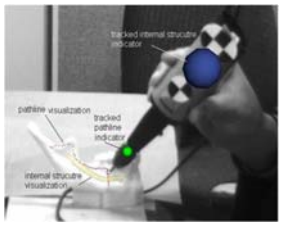

Figure 9: Totally augmented feedforward.

Feedback

Feedback should convey what the end user has already performed with respect to reaching the final goal. Regarding gesture recognition, the feedback can be simply ensured by plotting the gesture, although this process is subject to the lag problem [10], which may impediment the perception of immediate feedback. Feedback can be ensured by denoting the starting point of the gesture (e.g., initial point, plane, or volume), emphasizing salient portions of the gesture (e.g., subject to special attention) while de-emphasizing portions that are no longer of central attention.

Feedforward

Feedforward should convey what the end user should still perform with respect to reacting the initial goal by presenting any information relevant to this goal. For simple and medium gestures, feedforward can be ensured by representing the potential full gestures that can be finished while the end user is issuing one [3,5]. It therefore provides the end user with guidance on how the currently being issued gesture can be terminated in one or many directions. Information presented in the feedup component can also be here subject to more guidance by presenting to what extent these data are on the right track or to what extent they deviate from reference value (e.g., beyond a certain threshold) or reference interval (e.g., entering or leaving a bounded interval). For example, the average velocity can be displayed by suggesting the user to slow down the instantaneous velocity to increase precision, the instantaneous curvature can indicate a certain deviation with respect to a reference value. Special events, such as entering a zone, leaving a zone, pointing at to a certain element, can also be denoted.

Feedforward Experiment

For the purpose of the feedforward component, a controlled experiment was conducted in the context of maxilla-facial surgery to compare three feedforward scenarios with respect to a scenario without any feedforward serving as the baseline condition [12]:

1. No feedforward scenario (Fig. 6): the surgeon should cut a mandible by issuing a 3D gesture without any feedforward, but with feedback on the current gesture already issued so far.
2. A virtual feedforward scenario (Fig. 7): when the surgeon adopts the right gesture, a green sphere is presented on a separate screen to denote that the surgeon is on the right curve. Another sphere indicates the distance from the pointer tip to the internal structure as a continuous feedup.
3. A partially augmented feedforward scenario with gesture visualization and distance indicator with respect to the reference curve (Fig. 8): the feedforward is similar to Fig. 7 but presented on the real object by augmented reality and the sphere becomes translucent when the gesture pointer is occluded by the object. The gesture curve is also presented in translucency in case of occlusion.
4. A totally augmented feedforward scenario with gesture and nerve visualization and distance indicator for both gesture and nerve (Fig. 9): the feedforward is similar to Fig. 8, but with a second sphere associated with three colors: grey means “go ahead with the gesture”, blue warns the surgeon when the gesture is less than 3mm from the reference gesture, and red means “a sensitive region, like a nerve, is intersected”.

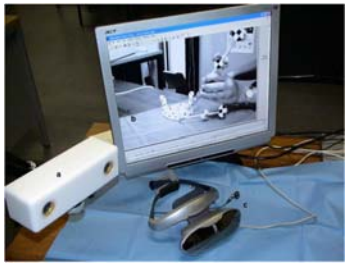

Figure 10: Design setup and apparatus.

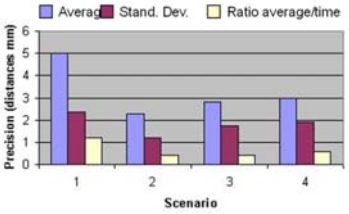

Figure 11: Measured precision.

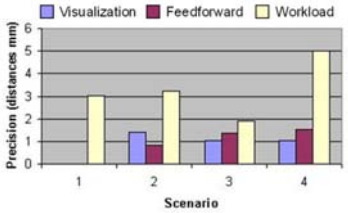

Figure 12: Perceived precision.

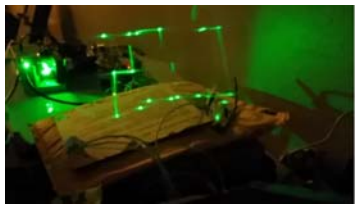

Figure 13: Holographic display.

Task. The goal of this experiment was to issue a complex curve 3D mid-air gesture that is as close as possible to the (ideal) reference gesture computed by the system before operation. The apparatus consists of a pair of augmented reality glasses coupled to a 3D object tracker with image fusion on a screen (Fig. 10).

Results and Discussion. Fig. 11 graphically depicts the precision measured as the average distance between the two gestures in the scenarios. The smaller the distance between gestures is, the better precision is obtained. The results suggest that the second scenario was the best in terms of average precision with its standard deviation and in terms of the ratio average/ time. Scenarios 3 and 4 produced larger distances than Scenario 2 both in average and in standard deviations, but the ratio remains comparable. Scenarios 2, 3, and 4 are all significantly more precise that Scen. 1.

Fig. 12 graphically depicts the perceived precision between the gestures in the four scenarios based on questionnaires and NASA-TLX forms regarding three variables: visualization, feedforward, and workload. Scenario 1 obviously has no visualization and no feedforward. The gain given by the visualization in scenarios 2-4 can be deduced from the assessed perception calibration task. We assume that the remaining performance gain is divided between user motor performance and feedforward. User motor performance was assessed during the motor calibration tests and we assume it as a constant for all scenarios. We found an average of 1.31mm with a standard deviation of 0.47 mm [12]. Regarding the workload, we tried to verify its relation with the visualization and feedforward variations. The highest workload rating coincide with the feedforward in Scenario 4, where the user has more information. Scenario 3 obtained the best workload.

3D Gesture Representations and Modalities

Representations. The aforementioned experiment involved only an indirect graphical gesture 2D representation. In order to involve a genuine 3D gesture representation instead, several devices are candidates:

- A *fog screen*, but the convex envelope in which the 3D gesture can be captured and represented is always large in length and height (e.g., 1 m), but not in depth (some cm.) and it only works as a vertical interaction surface, which may induce physical fatigue and deformation.
- A *volumetric display* may offer different types and sizes of a volume, but the 3D gesture will inevitably be physically constrained by the external surface of the device with limited depth (the gesture could be limited to only the immediate surface)
- A *holographic display* may offer a genuine 3D gesture representation that is subject to direct manipulation in space (e.g., Holovect, a holographic vector display capable of displaying images in 3D with light – Fig. 13), but manipulating gesture interaction points lacks precision. The touch hologram in mid-air [8] represents a better alternative for precision, but still requires augmented reality glasses.

Modalities. The above experiment also reveals that, although feedforward is delivered, the graphical channel becomes overloaded and could therefore benefit from a transposition into any other modality, such as sound, tactile, and haptic modalities. Here are two examples:

1. *Entering a 3D zone*: when a gesture is approaching a nerve zone, some feedup should indicate the probability of touching the nerve and deliver feedforward to avoid this zone. For instance, a vibration could be produced when approaching a sensible zone until a barrier is haptically felt and exceeded

after once the gesture has been pushed too far.

2. *Indicating an appropriate direction*: when a gesture produces a velocity towards an (in)appropriate direction, feedforward should warn the end user of this risk and suggest a correction while gesturing.

If the sound modality is used, sonic variables, such as location, loudness, pitch, timbre, rate of change, should convey feedup and feedforward. If the haptic modality is used, haptic variables, such as shape, hardness, temperature, weight, vibration, could be used.

Conclusion

This paper has introduced a decomposition of the feedback problem into three components, i.e. feedup, feedback, and feedforward, for curve 3D mid-air gestures. Gestures were only single-path (one path at a time) and single-stroke (no multistroke and therefore no segmentation problem). Future work will investigate how this same problem can be solved for multi-stroke gestures and different gestures in the same vocabulary.

Acknowledgements

This paper was supported by project FSR16 1C.11000. 023 financed by UCL, Belgium (2016-2019).

References

1. G. Bailly, E. Lecolinet, and L. Nigay. 2008. Flower Menus: A New Type of Marking Menus with Large Menu Breadth, Within Groups and Efficient Expert Mode Memorization. In *Proc. of AVI '08*, 15–22.
2. C. Bach, D. L. Scapin. 2003. Ergonomic criteria adapted to human virtual environment interaction. In *Proc. of IHM '03*. ACM Press, New York, 24–31.
3. O. Bau and W.E. Mackay. 2008. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. In *Proc. of UIST '08*. ACM Press, NY, 37–46.
4. F. Beuvens and J. Vanderdonckt. 2012. Designing Graphical User Interfaces Integrating Gestures. In *Proc. of SIGDOC '12*. ACM, NY, USA, 313–322.
5. W. Delamare, T. Janssoone, C. Coutrix, and L. Nigay. 2016. Designing 3D Gesture Guidance: Visual Feedback and Feedforward Design Options. In *Proc. of AVI '16*. ACM Press, New York, NY, 152–159.
6. J. Hattie and H. Timperley. 2007. The power of feedback. *Review of Educ. Research 77*, 1, 81–112.
7. A. Kamal, Y. Li, and E. Lank. 2014. Teaching motion gestures via recognizer feedback. In Proc. of IUI '14. ACM Press, New York, NY, 73–82.
8. C. Kervégant, F. Raymond, D. Graeff, and J. Castet. 2017. Touch hologram in mid-air. In *Proc. of SIGGRAPH Emerging Tech. 2017*. 23:1–23:2
9. B. Long, S. A. Seah, T. Carter, and S. Subramanian. 2014. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Transactions on Graphics 33*, 6: 181:1-181:10.
10. U. Braga Sangiorgi, V. Genaro Motti, F. Beuvens, and J. Vanderdonckt. 2012. Assessing lag perception in electronic sketching. In *Proc. of NordiCHI 2012*. ACM Press, New York, NY, 153–161.
11. R. Sodhi, M. Glisson, and I. Poupyrev. 2013. AIREAL: tactile gaming experiences in free air. In *Proc. of SIGGRAPH Emerging Technologies 2013*. 2:1.
12. D. Trevisan, J. Vanderdonckt, B. Macq, and Ch. Raftopoulos. 2003. Modeling interaction for image-guided procedures. In *Proc. of Medical Imaging: Image-Guided Procedures 2003.*
13. J. Vanderdonckt. 2010. Distributed user interfaces: how to distribute user interface elements across users, platforms, and environments. *Inter.'10*.
14. R.-D. Vatavu. 2013. The impact of motion dimensionality and bit cardinality on the design of 3D gesture recognizers. *Int. J. Human-Computer Studies 71*, 2013, 387–409.
15. J. Vermeulen, K. Luyten, E. van den Hoven, and K. Coninx. 2013. Crossing the bridge over Norman's gulf of execution: revealing feedforward's true identity. In *Proc. of CHI '13*. ACM, NY, 1931–1940.
16. G.A. Wilson and S.A. Brewster. 2017. Multi-moji: Combining Thermal, Vibrotactile & Visual Stimuli to Expand the Affective Range of Feedback. In Proc of CHI '17. ACM Press, New York, NY, 1743–1755.