

Master Creative Intelligence & Technology

Distortion as Acceptance: A Tangible Audio Interface for Experiencing Negative Self-Referential Thoughts

Name	Sabine Perigault Monte
Student ID	1847392
Date	06/10/2026

1st supervisor	Max van Duijn
2nd supervisors	Alexandra Blank, Edwin van der Heide

Master's Thesis

Leiden Institute of Advanced Computer Science
Leiden University
Niels Bohrweg 55
2333 CA Leiden
The Netherlands

Distortion as Acceptance: A Tangible Audio Interface for Experiencing Negative Self-Referential Thoughts

Sabine Perigault Monte

MSc Thesis, Creative Intelligence and Technology, Leiden University
Primary Supervisor: Max van Duijn
Secondary Supervisor: Alexandra Blank
Secondary Supervisor: Edwin van der Heide October 2026

ABSTRACT

Almost everyone experiences negative self-referential thoughts. For some people, these thoughts shape self-image and behaviour: they tend to be believed, which makes them hard to let go of. Cognitive defusion within Acceptance and Commitment Therapy changes the relationship people have with their thoughts. Existing defusion techniques have people say the thought aloud, for example by repeating it rapidly or voicing it in an exaggerated tone. What remains unexplored is the effect of hearing one's own recorded thought and changing how it sounds. This study examined whether this influences the impact of the thought. Eighteen participants recorded their own negative self-referential thought and altered it with a tangible audio interface offering four modulations: pitch, robot, reverb and panning. Believability, discomfort and acceptance were measured before and after as indicators of defusion. Across the whole interaction, believability and discomfort decreased significantly, while the change in acceptance was not significant. Robot gave the most consistent reductions in believability and increases in acceptance, and reverb increased discomfort. Pitch produced no significant effect, but the direction of change varied between participants. Panning showed no significant effect. Several participants showed both increased and decreased intensity within the same session. There was no baseline audio condition. The study is exploratory, therefore the results do not confirm a hypothesis. Results rather show an indication of a pattern.

INTRODUCTION: NEGATIVE SELF-REFERENTIAL THOUGHTS

Negative self-referential thoughts are part of the broader concept of self-talk. Self-talk refers to the internal verbal dialogue we have with ourselves, which is commonly known as our "inner voice." These self-directed expressions can take many forms: positive or negative, instructional or motivational, and may occur either silently or out loud [41, 9]. Self-talk serves a number of functions, including impulse inhibition, regulating courses of action and emotion regulation. Self-referential thoughts refer to thoughts that take the self as their object. These are claims or evaluations about who one is, not about a situation, an event, or another person. When negative, these thoughts tend to evaluate the self as a whole. This could explain why they carry more emotional impact. Negative self-referential processing is seen as a risk factor for the onset and recurrence of depressive episodes [30, 35]. It has also been linked to a more general disposition towards negative emotionality and distress [48]. Therefore, these thoughts could influence a person's overall sense of identity and worth.

The present research focuses on a specific form of self-talk: negative self-referential statements that are critical, derogatory, or self-defeating, such as "I am not good enough," "I am worthless," or "My life is pointless". Understanding why these thoughts are so difficult to disengage from requires a closer look at the psychological process of cognitive fusion, explored in the next section: 'Why does it feel real?'

COGNITIVE FUSION: WHY DOES IT FEEL REAL?

Acceptance and Commitment Therapy (ACT) is part of a broader family of cognitive behavioural therapies (CBT). ACT aims to build a more flexible relationship with internal experiences. Traditional cognitive behavioural methods, such as cognitive restructuring, aim to evaluate and modify the content of negative thoughts by questioning whether they are accurate or helpful. ACT instead works on the individual's relationship with those thoughts, so they can function effectively regardless of what their inner voice is saying [20]. That relationship, however, does not exist in isolation. It always takes place within a psychological and social context. It is precisely this context that determines the degree to which verbal events, such as thoughts and rules, regulate behaviour [20, 21]. ACT is grounded in functional contextualism, a framework that evaluates psychological events by their function within a given context, not by their form or content. Alongside this, ACT draws on Relational Frame Theory (RFT), a behavioural account of human language and cognition that explains how humans learn to relate events to one another through language, and how these learned verbal relations can come to govern behaviour [4]. RFT explains why a person can be emotionally triggered independent of the circumstances. The relational framing allows people to infer a relation between events without a direct experience. A single past failure will be verbally linked to a negative self-evaluation (for instance; 'I am incompetent') and then projected onto an unrelated future scenario ('I will fail this presentation too'). Together, these two frameworks explain why thoughts, even when unhelpful or inaccurate, can wield such a strong influence over how a person feels and acts. Private events, including thoughts, sensations and

emotions, are therefore approached by targeting their function, rather than trying to alter their form or frequency.

Psychological flexibility is a central concept in ACT, defined as the ability to contact the present moment and to persist in or change behaviour in the service of chosen values [19]. If we take the previous example, this would mean that even though one thinks they would fail the presentation, they would still present, not because the fear is gone, but because it aligns with something that they value. Psychological inflexibility, its counterpart, arises when internal experiences such as thoughts, emotions, and memories dominate behaviour in ways that are unhelpful. The ACT model (see Figure 1) identifies six core processes that together contribute to psychological flexibility, each with both a pathological and a therapeutic counterpart. These describe what goes wrong and what can change. Cognitive defusion is one of these six processes, with cognitive fusion representing its pathological counterpart. The process of cognitive defusion will be explained in the next section: 'What can change?'

Before turning to defusion, this section first looks at its counterpart. Cognitive fusion refers to the tendency to become so entangled with one's own thoughts that they are taken to be literally true. A passing mental event no longer registers as such, the individual responds to the thought as though it were a direct and literal reflection of who they are or how the world is [20, 18]. This process is rooted in how language works. Through what is known as verbal processes, the mind constantly generates descriptions, evaluations, and rules about the self and the world [20]. These verbal events, including thoughts such as "I am not good enough," are typically treated as though they correspond directly to reality, a phenomenon referred to as a context of literality. In this context, the mind does not automatically distinguish between a thought and a fact, causing a self-critical statement to stop feeling like an interpretation and start feeling like a truth. From an ACT perspective, the harm lies in the context in which a thought is treated as literal truth, and in the behavioural regulation that follows from it, not in the thought itself. The more fused an individual is with a thought, the more they believe it, and the more it governs their emotional experience and behaviour.

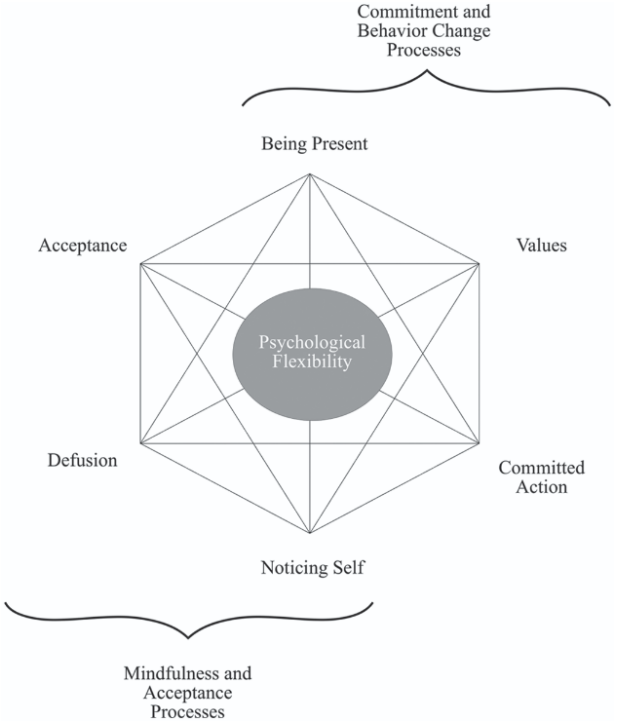


Figure 1: The ACT hexaflex model, illustrating the six core processes of psychological flexibility.

COGNITIVE DEFUSION: WHAT CAN CHANGE?

Cognitive defusion is conceptualised as altering the literal meaning and behaviour-regulatory function of private events, without necessarily changing their form, frequency, or situational sensitivity [7]. Hayes refers to this process as deliteralisation, learning to hold thoughts less literally rather than eliminating them [20]. This is why ACT places such emphasis on the role of language. According to Relational Frame Theory, verbal relations are bidirectional. The words "I am not good enough" can evoke the same psychological response as if the statement were literally true, simply because the mind has learned to treat the verbal event and the experience as equivalent [14]. ACT does not regard language as inherently problematic, as it serves many important functions in human life. However, ACT takes a pragmatic stance towards the relationship individuals have with language, recognising that the rigid, literal mode in which language is often used can become problematic when private experiences are treated as though they correspond directly to an objective reality. When the same evaluative language continues to govern behaviour without producing meaningful change, defusion offers a way to relate to that language differently, without needing to abandon it.

Simply instructing individuals to take their thoughts less literally is rarely effective on its own. That is why experiential learning is such an important factor. By drawing on metaphors and defusion techniques, they aim to create a different relationship with language from the inside out, beyond what analytical instruction alone can achieve [47]. The use of metaphor is central to this approach. Defusion can be understood as creating a space between the individual and their internal experiences, allowing thoughts to be observed without being acted upon [6].

Defusion techniques take many forms in practice. Some work through language, such as prefacing a thought with "I am having the thought that…" to create linguistic distance. Others rely on visualisation, repetition, or embodied experience to alter the relationship with a thought [7, 50]. What these approaches share is an emphasis on experiential change over analytical understanding: the goal is not to think differently about a thought, but to experience it differently. The following section, "From concept to interface," describes how these principles informed the design of the present research.

FROM CONCEPT TO INTERFACE

The following section describes the design rationale and the theoretical choices that informed the experiment and interface. The full research question is outlined at the end of this section.

The interface used in this research consisted of a microphone, a sampler, and a screen displaying the controls and instructions. Participants recorded their own negative thought, which was then played back and manipulated through a set of physical controls, each corresponding to a different audio modulation. The conceptual design of the interface was grounded in the Radio Doom & Gloom metaphor from ACT [20], which describes the mind as continuously broadcasting a stream of thoughts that cannot simply be switched off. Within the metaphor, turning off the radio is not an option, but changing how one relates to what one hears is. The interface invited participants to change how the thought sounds. Although the interface did not take the physical form of a radio, this metaphor shaped its design. Within ACT, metaphors are used to bypass analytical language in favour of direct experience. This aligns with Conceptual Metaphor Theory, which suggests that abstract concepts are more effectively understood when mapped onto concrete, familiar experience [29]. Embodied cognition further supports this as understanding is rooted in physical and sensorimotor experience [11, 32]. A tangible interface offers a form of engagement that goes beyond reading or listening. Because participants can actively modulate the sound of their own thought, the interface may elicit defusion through physical action.

Several established defusion techniques informed the design of this research. These techniques share a common principle: each works by altering the language parameters that give a thought its meaning and emotional weight. This depends in part on the conventions through which language is delivered; when these are disrupted, the perceived credibility and emotional impact of a thought can weaken [7].

The word repetition technique, first described by Titchener [49], draws on the phenomenon of semantic satiation. Under normal circumstances, words automatically activate their associated meanings and emotional responses. However, when a word is repeated continuously, this connection weakens. Sustained exposure to the sound of a word causes its semantic content to fade into the background, leaving only the auditory experience [45]. Empirical support for this technique in the context of negative self-referential thoughts was provided by Masuda et al. [34], who demonstrated that rapid vocal repetition of a one-word version of a negative self-referential thought significantly reduced both its perceived believability and emotional discomfort.

A related technique targets the same connection through a different mechanism. Instead of relying on prolonged exposure, it disrupts the familiar vocal qualities that give a thought its emotional weight by altering how the thought is spoken. This can be done, for example, by saying it in a slow, exaggerated voice, singing it aloud, or delivering it in a silly tone [7].

The audio interface essentially integrates components of these two techniques described above (word repetition and style of speech) in the experiment. Both of these techniques involve the participant actively vocalising the thought. In this experiment, however, the interface adapts these techniques in a different way. Participants record their thought once and subsequently listen to it while actively modulating their own voice. Is there anything beneficial to listening to one's own voice? One study found that one's own voice undergoes unique neural processing that distinguishes it from the voices of others. This processing engages self-referential networks, which may be relevant for self-talk as an emotion regulation strategy. In this study, the researchers compared cognitive defusion and self-affirmation as emotion regulation strategies, and found that defusion more strongly activated self-referential networks when the own voice was used. [23]. Negative self-referential thoughts are inherently personal, so participants' own voices were used here, as they were considered more likely to engage the self-referential processing that characterises the original experience of the thought.

However, have you ever listened to your own voice from a tape or recording? People typically perceive their recorded voice differently from their actively spoken voice. A recording can produce a sense of unfamiliarity or discomfort. One study investigating this phenomenon found large individual differences in how representative a recorded voice feels, and concluded that there is no universal filter that can make a recorded voice sound identical to the voice one hears while speaking [26]. In the context of the present research, this limitation was acknowledged but not corrected for, as the audio modulations applied to the recordings inherently altered their vocal properties.

Pitch & Timbre

Pitch and timbral modulations were incorporated as mechanisms through which the acoustic properties of a participant's recorded thought could be altered. Research on vocal emotion perception suggests that fundamental frequency (F0), which largely determines how high or low a voice sounds, and timbre are the primary parameters through which emotional states are communicated in speech [40, 24]. Both pitch alteration and timbral modulation target these acoustic properties. Manipulating them disrupts the cues that give a thought its emotional quality, possibly changing how it is perceived. There is however a difference between F0 and timbre. The pitch modulation used here combines a pitch shift with a formant correction, resulting in a more natural-sounding high or low voice, in contrast to a pitch shift alone, which distorts the formants and produces an unnatural sound. The robot modulation alters the timbre of the voice through a resonant filter, this results in a more unnatural and mechanical sound. Whether this more unnatural quality disrupts the emotional cues differently than pitch is explored in the results.

Reverb & Panning

Reverberation and panning were incorporated to target a different dimension of the recorded thought, its perceived spatial location. The spatial properties of sound, including reverberation and stereo positioning, have been shown to influence

listeners' emotional responses and engagement [46]. By placing the thought within a different acoustic space through reverberation, or moving it between audio channels through panning, the perceived spatial origin of the thought is altered. This spatial modulation was chosen to create a sense of distance between the participant and the recorded thought, which supports the defusion process. Volume was included in the interface but not incorporated in the quantitative analysis. A pilot experiment suggested it produced no notable effect, and including it would have extended the measurement burden considerably.

How these manipulations are experienced varies between individuals and is explored further in the results section.

The decision to combine multiple defusion techniques within a single interface followed naturally from the radio metaphor that informed the design. The looping of the recorded voice incorporates the word repetition technique, while the audio modulations allow participants to alter their voice. Separating these elements would not reflect how the interface is actually used.

Together, these theoretical and design choices informed the central research question of this study: To what extent does interacting with a metaphor-based tangible audio interface facilitate cognitive defusion and reduce the impact of negative self-referential thoughts in participants?

METHOD

Participant and setting

The study was conducted at Leiden University. Eighteen participants were recruited through social media, personal networks and university communications. No formal demographic data was collected. Given the sensitive nature of the experiment, the privacy of participants was prioritised. As the study was exploratory, the focus was on whether the intervention produced any effect, rather than on demographic differences. Participants were asked to report any hearing impairment or tinnitus. Those with severe hearing loss were excluded, as the task required adequate audio perception. A warning regarding the potential impact of audio manipulations on tinnitus was included in the consent form. The experiment was conducted in both Dutch and English, with participants completing all materials in their preferred language.

A small sample size was anticipated, therefore, a within-subjects design was employed. Dividing participants into an experimental and control group was not feasible, and a pre-post comparison within the same group was considered the most appropriate approach. Participants were exposed to a single experimental condition, namely interaction with the audio interface. Believability, discomfort and acceptance were measured pre and post experiment. Before the experiment, a CFQ-13 was administered. Post experiment, participants had to complete a questionnaire consisting of Likert-scale items and open-ended questions.

Privacy

To protect the privacy and create conditions in which a personal negative thought could be used without disclosure, several measures were taken. The thought remained private throughout the experiment, participants did not need to share their thought (only if they chose to). Prior to recording, the headphone setup was demonstrated using a demo recording to confirm the thought would not be audible to anyone but the participant. During the recording and the interface interaction, the researcher left the room. After the experiment, the audiofile was deleted in front of the participant. As ACT aims to change the functional effects of a thought rather than its content, protecting the privacy was both ethically and theoretically motivated.

Procedure

Each experiment session took approximately 40–50 minutes including consent & debriefing. No fixed time limit was imposed in order to prevent any feeling of performance pressure. There were no sessions that exceeded one hour. Participants were not informed about the theoretical background of the interface or its connection to cognitive defusion techniques. To ensure all participants received identical information prior to the task, instructions were presented in written form rather than verbally. The instruction text is included in Appendix A.

Thought selection

Participants were asked to identify one negative self-referential thought that they experienced repeatedly and found personally distressing and believable. To support this process, a worksheet was provided to help participants formulate and articulate their thought. The worksheet is included in Appendix C (English) and Appendix D (Dutch). At no point were participants required to share their negative self-referential thought with the researcher, it remained private throughout the experiment. The thought selection was conducted by the researcher, who guided participants through the thought selection process using a standardised set of instructions, see Appendix B. A script was written in Dutch and English, so each session the same introduction and explanation was used.

Pre-measures: VAS Believability, Discomfort and Acceptance

The impact of the negative self-referential thought was assessed across three dimensions: believability, discomfort and acceptance. This was assessed using a visual analogue scale before and after the experiment. Responses ranged from 0–100 with 0 (not uncomfortable at all) to 100 (very uncomfortable) for the emotional discomfort, 0 (not believable at all) to 100 (very believable) for the believability scale, and 0 (not accepted at all) to 100 (fully accepted) for the acceptance scale. Believability and emotional discomfort were selected as outcome measures following Masuda et al. [34, 33], who used these dimensions to measure the effect of a cognitive defusion technique. In the present study the same outcome measures are used to assess the effect of the interface on the functional impact of the thought, following the same operationalisation of cognitive defusion as Masuda et al. Acceptance was additionally included to examine whether the interface increased participants' comfort with having the thought present, by asking how okay it was for them that the thought was present before and after the interaction.

Pre-measures: CFQ-13

Cognitive fusion was measured prior to the experiment using the Cognitive Fusion Questionnaire [16], a 13-item self-report measure with total scores ranging from 13 to 91, with each item rated on a 7-point scale from 1 (rarely/never) to 7 (always), where higher scores indicate greater cognitive fusion. The experiment aims to facilitate cognitive defusion, therefore, it was relevant to establish the degree to which participants were fused with their thought prior to the experiment. Defusion can be seen as both a technique and a skill. As a technique, it refers to a specific practice or intervention that reduces the impact of thoughts on behaviour and experience without necessarily changing their content [7], which can occur within a single interaction. As a skill, it refers to a more durable ability developed over time and therefore cannot be assessed in a single session. The defusion process itself is not observable; what is measured instead is its functional impact on the thought, through the three chosen outcome measures. This study aims to facilitate cognitive defusion through a technique (the interface), therefore the CFQ-13 was administered only once as a sample characteristic.

Recording

After completion of the survey, participants were instructed how to record their thought. A Novation Circuit Rhythm sampler [39] was used as the primary control device. Buttons were mapped to record, play and stop. The researcher demonstrated the recording procedure using a sample recording. When the participant was confident enough to record their own thought, the researcher left the room and came back as soon as the participant was done recording. Participants could record their thought up to three times and select the recording they preferred.

Graphical user interface

Once participants had finished their recording, the researcher launched the custom graphical user interface (GUI). A screen displayed the GUI, showing a slider for each audio modulation (pitch, robot, reverb, panning and volume). The sliders moved in real time in response to the physical controls on the Novation sampler, providing participants with visual feedback of the active modulations. A legend showing instructions for recording and resetting the audio modulations was positioned in the upper left corner of the screen. The GUI was intentionally designed to be plain and simple, without too many distractions, to keep participants focused on the audio rather than the screen. Participants were instructed to listen to each audio modulation in isolation before combining them (this instruction was also displayed on screen, see instruction text below), so that they could clearly identify what each modulation sounded like individually. This made it possible to assess the effect of each modulation individually as well as in combination. Once familiar with each effect, participants then freely explored and combined the modulations as they chose. This element of free exploration was deliberate as it reflects ACT's emphasis on experiential learning. Instead of conversing or imagining, the participants experienced the transformed sound of their own voice in real time. The audio was played in a loop, but participants were able to stop the recording if the experience was too uncomfortable or distressing. The researcher walked participants through the interface with the audio on mute, so that the sound of each modulation was encountered for the first time during the experiment.

Instruction text:

Pay attention to what you experience.

1: Go through each modulation separately first.

Press Reset if you want to return the function to its original value.

2: You can now freely experiment and combine everything.

Post-measures: VAS and Content Analysis

Following the interaction phase, participants completed a second VAS assessment (same as the VAS prior to the experiment). Each modulation was also evaluated separately, with participants rating its effect on believability, emotional discomfort and acceptance. This made it possible to compare the effects of each modulation against each other. The ratings were scaled from -50 to $+50$. For believability, a higher score indicated the thought felt more believable. For emotional discomfort, a higher score indicated more discomfort. For acceptance, a higher score indicated greater acceptance of the thought.

The post-session questionnaire consisted of six questions. Open and closed ended questions were included. This study is exploratory and aimed at the experience of modulating the audio of one's negative self-referential thought, therefore, open ended questions were added to capture the subjective experience of participants in their own words. Three open-ended questions addressed moments in which a shift occurred in their experience of the thought, which combinations of modulations contributed to this, and whether there were moments in which no shift occurred.

Open-ended questions were analysed via a qualitative content analysis [12]. Content analysis was chosen for its ability to systematically categorise language through both deductive and inductive approaches, and for its suitability in exploring participants' subjective experiences of the interaction. The categorisation matrix was refined iteratively during coding, conducted using Taguette [42], consistent with standard practice in qualitative content analysis [44]. The full categorisation matrix, including category definitions and illustrative examples, is available as a codebook. See Appendix E.

Categories for experiential shifts were derived from theoretical frameworks relevant to cognitive defusion [7].

The following categories were used to code participant responses:

Experiential shifts:

- Increased intensity: words and phrases describing a stronger experience of the sentence
- Decreased intensity: words and phrases describing a weaker experience of the sentence

- Outsider perspective: viewing the sentence from an external position
- Insider perspective: experiencing the sentence from a first-person viewpoint
- Deliteralisation: the sentence felt less meaningful or less real
- No shift: explicit indication of no change in experience
- Descriptive: change described without clear direction

Participant responses were additionally coded for metaphorical language, capturing the figurative expressions used to describe the experience of the audio modulations. Categories were developed inductively from the data, and where responses contained language corresponding to multiple categories simultaneously, all relevant codes were assigned.

- Spatial: metaphors of location or distance, describing where the thought was perceived to be.
- Personification: attributing a separate character, voice or intention to the thought
- Weight: using weight or heaviness to describe the impact of the thought.
- Sound texture: metaphors of tonal quality or texture, reflecting how the modulated audio was heard.
- Atmospheric: metaphors of mood or atmosphere evoked by the modulation, reflecting how the experience felt.
- Repetition: metaphors describing the experience of hearing the thought on repeat.

Affective responses were coded based on emotional states explicitly reported by participants, as well as affective states inferred by the researcher from the language used, following the principles of emotion coding as described by Saldaña [44].

The following categories were identified:

- Discomfort
- Humour
- Hurt
- Loneliness
- Pleasantness
- Sadness
- Seriousness
- Self-compassion

Materials and Setup

The materials consisted of a Shure microphone, a MOTU M6 audio interface, a Novation Circuit Rhythm sampler, headphones, and two screens. One screen displayed the session instructions, rendered through a custom Python script. The second screen, connected to the researcher's laptop, displayed the Ableton Live interface used to manage the audio modulations. See Figure 2.

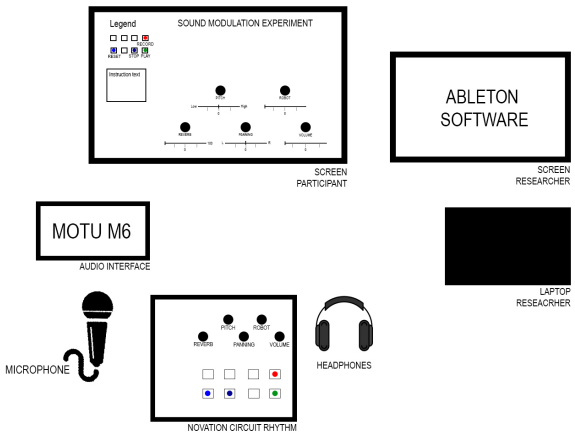


Figure 2: Setup of the experiment.

Ableton

In Ableton [1] an audio effect rack was created, an audio rack is a chain of audio effects/devices that are used for the audio modulation. The chain starts with a compressor for volume control. After the compressor, the Graillon 3 (a VST) [2] is used for shifting the pitch (from high to low). The robot modulation was achieved using Swirl, an audio filter native to Ableton that applies a resonant filter effect to the voice. The reverb modulation used Outer Space, a preset that simulates how sound reflects and decays in a large physical space. Utility is used to control the volume, the gain was adjusted to have a minimal dB of -20 and a max of 35 dB. Participants could not turn the volume down completely, ensuring continuous audio playback throughout the experiment. For the panning of the audio the E4L source panner [13] was used. In this setup played through standard stereo headphones, the perceptible effect resembles conventional left-right panning, but the E4L source panner is designed for 3D (Ambisonic) spatial positioning so there was a subtle spatial depth for the attentive listener. This was chosen deliberately because the ambisonic architecture of the E4L plugin could be a foundation for future work involving more advanced spatial movement of sound. Limiter was used to close the audio chain preventing digital clipping and distortion.

Graillon, Swirl, Outer Space, Utility and the E4L Source panner were mapped via Ableton's MIDI Map Mode to specific knobs on the Novation Circuit Rhythm for real-time manipula-

tion of the modulation parameters during the experiment (see Table 1). The settings of each device are shown in Figure 3.

Table 1: Mapping of the macros to the devices and parameters in Ableton.

Macro	Device & parameter
Pitch Shift	Graillon 3 – Pitch shift
Reverb	Outer Space – Dry/Wet
Robot	Swirl – Dry/Wet
Volume	Utility – Gain
Panning	E4L Source Panner – Azim

RESULTS

CFQ-13

The CFQ-13 pre-scores showed a mean cognitive fusion score of $M = 52.11$ ($SD = 11.46$).

Prior to analysis, the assumption of normality was assessed using the Shapiro–Wilk test. Results indicated that the data did not significantly violate normality for the paired samples analyses (all ps $> .05$). However, given the small sample size ($N = 18$), which limits the power of normality tests to detect violations, non-parametric analyses were used as a conservative approach. Wilcoxon signed-rank tests were conducted to examine pre- to post-intervention changes across three VAS measures ($N = 18$). No Bonferroni correction was applied. Results are reported at $\alpha = .05$ and should be interpreted as indicative rather than confirmatory. A sensitivity analysis (G*Power 3.1.9.4) indicated that, with $N = 18$, the study had 80% power to detect effects of $d_z \geq 0.72$ (two-tailed). To compare the observed effects with this threshold, effect sizes (d_z) were additionally calculated from the corresponding paired-samples and one-sample t-tests, as G*Power expresses the sensitivity threshold for the Wilcoxon signed-rank test in terms of d_z. For the modulation ratings, one-sample Wilcoxon signed-rank tests were conducted against the neutral midpoint. The same sensitivity threshold ($d_z \geq 0.72$) reported above applies to these analyses, given the identical sample size and test parameters.

Believability decreased significantly from pre- ($M = 82.28$, $SD = 11.07$) to post-intervention ($M = 67.11$, $SD = 24.70$), $p = .005$, $r = 0.809$, indicating a large effect.

Discomfort decreased significantly from pre- ($M = 59.00$, $SD = 25.08$) to post-intervention ($M = 49.78$, $SD = 26.94$), $p = .017$, $r = 0.708$, indicating a large effect.

No significant change was found for **acceptance** from pre- ($M = 71.78$, $SD = 21.09$) to post-intervention ($M = 77.39$, $SD = 20.74$), $p = .221$, $r = -0.367$, though the effect size was medium. A negative r value here reflects that post-intervention scores were higher than pre-intervention scores, consistent with the hypothesised increase in acceptance.

Spearman correlations

Spearman correlations were conducted to assess whether cognitive fusion scores (CFQ-13) correlated with change scores on the three VAS measures. No significant correlations were found, suggesting that pre-intervention levels of cognitive fusion did not predict the degree of change in believability, discomfort or acceptance. A significant positive correlation was found between changes in believability and discomfort ($\rho = 0.491$, $p = .039$; see Figure 4), indicating that participants who reported greater reductions in believability also tended to report greater reductions in discomfort. This is consistent with Masuda et al. [33], who identified emotional discomfort and believability as related aspects of the psychological reaction to negative self-referential thoughts.

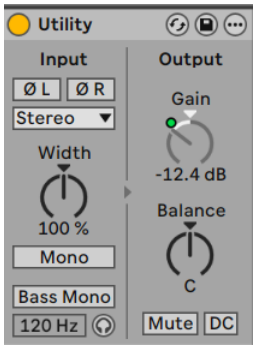
Quantitative

Pitch

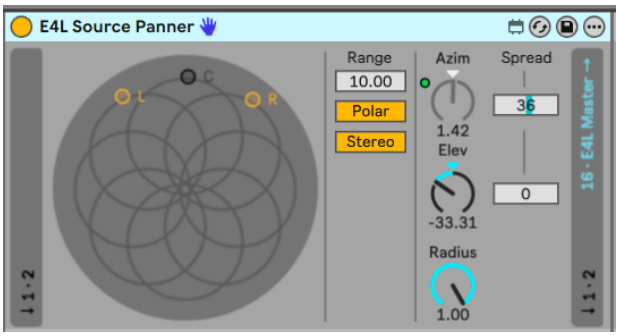
Analysis of change scores revealed no significant effects for pitch across any of the three outcome measures (believability: $V = 87$, $p = .965$, $r = 0.018$; discomfort: $V = 60.5$, $p = .717$, $r = -0.110$; acceptance: $V = 71.5$, $p = .531$, $r = 0.192$). This likely reflects individual variability in the direction of effects rather than a uniform absence of effect: for believability, scores ranged from -43 to 45 ($M = 1.67$, $SD = 32.11$), with 10 of 18 participants rating the thought as more believable and 8 as less believable. The raincloud plots illustrate this variability (Figure 5).



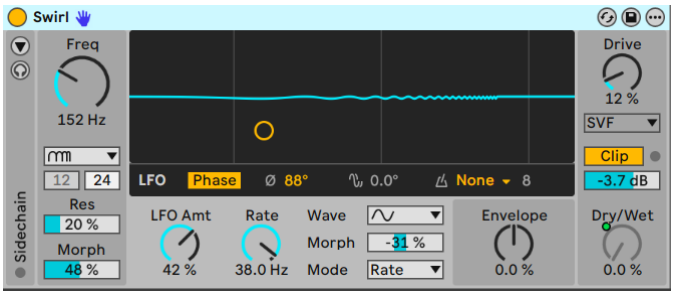
(a) Graillon 3



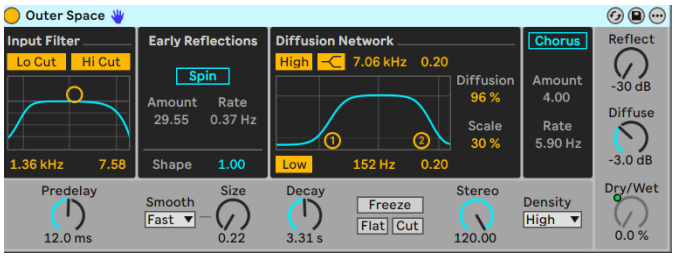
(b) Utility



(c) E4L Source Panner



(d) Swirl



(e) Outer Space

Figure 3: Settings of the audio devices in the Ableton effect rack.

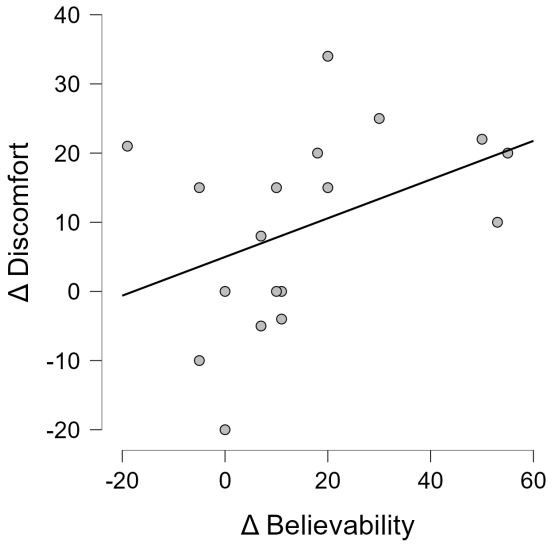
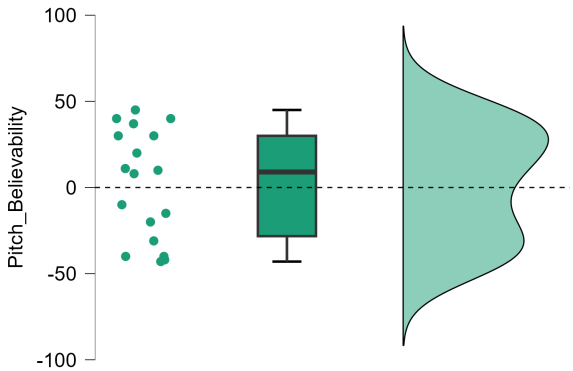
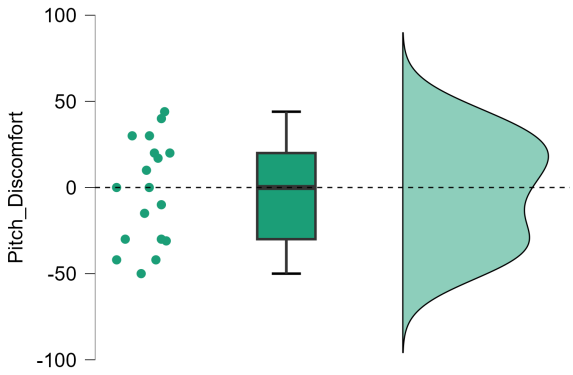


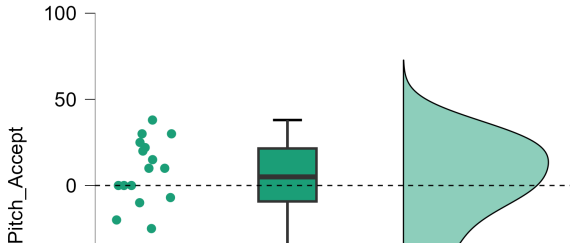
Figure 4: Correlation between change in believability and change in discomfort ($\rho = 0.491$, $p = .039$).



(a) Believability

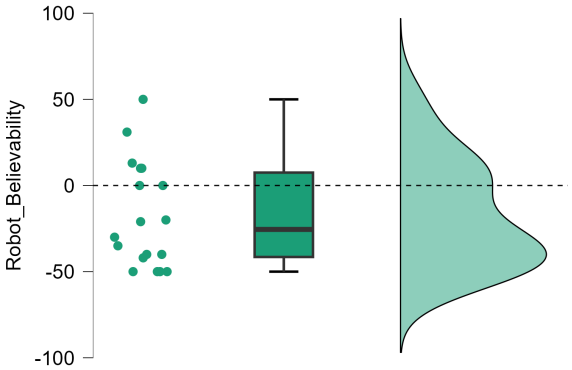


(b) Discomfort

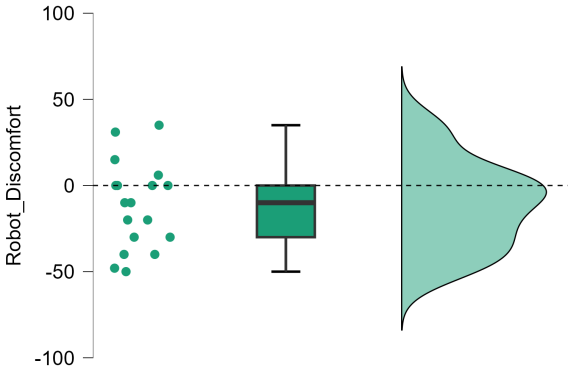


Robot

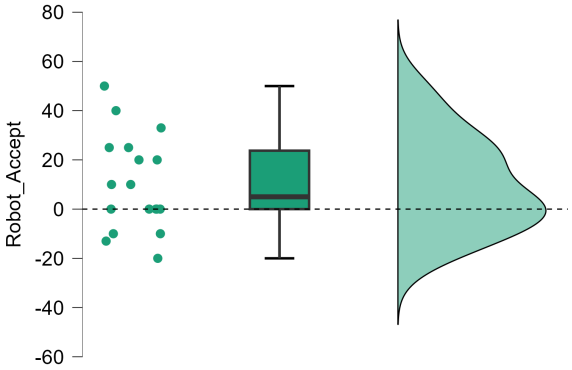
Analysis of change scores revealed significant effects for robot on believability ($V = 27$, $p = .036$, $r = -0.603$) and acceptance ($V = 74$, $p = .049$, $r = 0.626$), indicating that the majority of participants rated the thought as less believable and more acceptable following the robot modulation. Discomfort did not reach significance ($V = 24$, $p = .078$, $r = -0.543$), although the majority of participants rated the thought as less distressing following the robot modulation, with a small number of participants showing an increase (Figure 6).



(a) Believability



(b) Discomfort

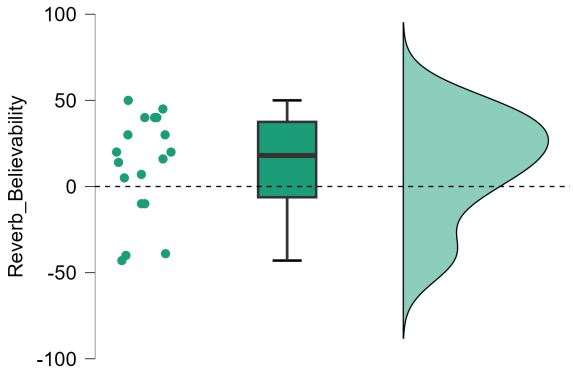


(c) Acceptance

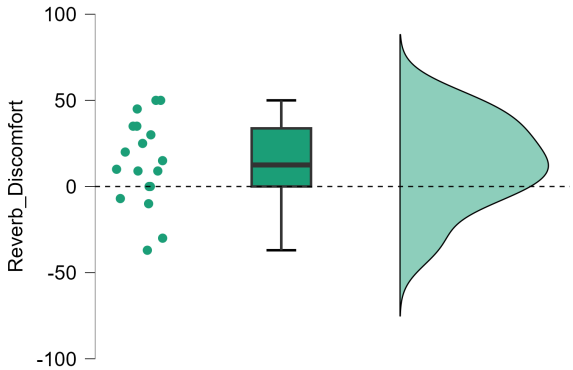
Figure 6: Raincloud plots of the change scores for robot.

Reverb

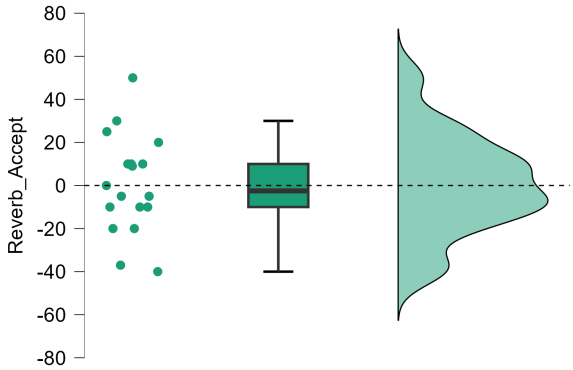
Analysis of change scores revealed a significant effect for reverb on discomfort ($V = 108$, $p = .041$, $r = 0.588$), but in the opposite direction. The score indicates that the majority of participants rated the thought as more distressing following the reverb modulation. Effects for believability did not reach significance ($V = 123.5$, $p = .102$, $r = 0.444$), although the majority of participants rated the thought as more believable following the reverb modulation (Figure 7). The distribution shows a wide spread, with a small number of participants showing a decrease. No effect was found for acceptance ($V = 77.5$, $p = .981$, $r = 0.013$).



(a) Believability



(b) Discomfort

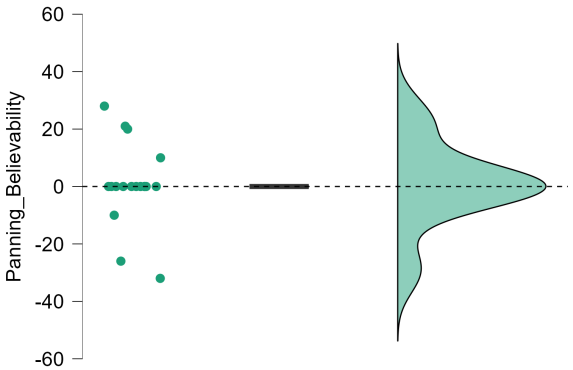


(c) Acceptance

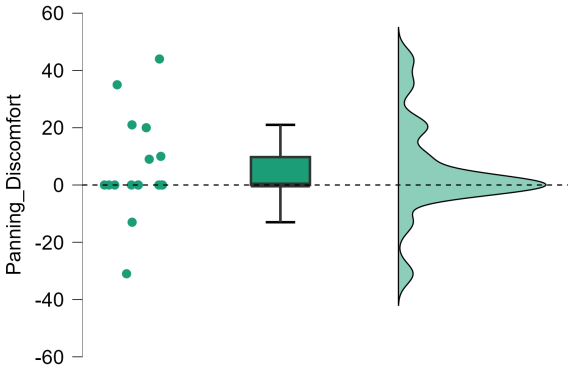
Figure 7: Raincloud plots of the change scores for reverb.

Panning

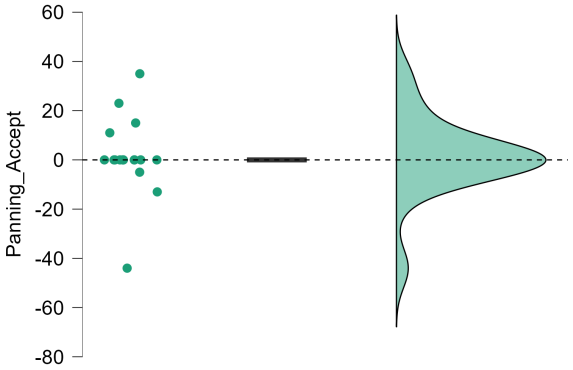
Analysis of change scores revealed no significant effects for panning across any of the three outcome measures (believability: $V = 14.5$, $p = 1.000$, $r = 0.036$; discomfort: $V = 27$, $p = .234$, $r = 0.500$; acceptance: $V = 17$, $p = .673$, $r = 0.214$). The raincloud plots show that scores for believability and acceptance clustered around zero, suggesting that panning had little to no effect on these outcomes for most participants (Figure 8).



(a) Believability



(b) Discomfort



(c) Acceptance

Figure 8: Raincloud plots of the change scores for panning.

Qualitative: content analysis

After using each audio modulation individually, participants then entered a free exploration phase, during which they could combine the audio modulations in any chosen way. Pitch combined with reverb was the most frequently reported combination ($n = 6$), followed by pitch combined with robot and reverb ($n = 3$) and pitch combined with robot ($n = 3$). In every reported combination, either pitch or robot was present (see Table 2).

Table 2: Combinations of audio modulations reported during the free exploration phase.

Combination	*n*
Pitch + Reverb	6
Pitch + Robot + Reverb	3
Pitch + Robot	3
Pitch + Volume + Reverb	2
Pitch + Volume	1
Panning + Pitch + Reverb	1
Robot + Reverb	1

The results of the qualitative content analysis for experiential shifts, metaphors and affective state indicators are presented in Table 3. An overview per participant is included in Appendix E.3. For all the definitions of the categories and general coding rules, see the codebook in Appendix E.

Table 3: Number of participants per category ($N = 18$).

Category	*n*
Experiential shifts	
Increased intensity	5
Decreased intensity	4
Outsider perspective	4
Insider perspective	2
Deliteralisation	2
No shift	2
Descriptive	1
Metaphors	
Spatial	6
Personification	4
Atmospheric	3
Weight	2
Sound texture	2
Repetition	2
Affective states	
Discomfort	4
Humour	4
Pleasantness	2
Seriousness	2
Hurt	1
Loneliness	1
Sadness	1
Self-compassion	1

Experiential shifts

All the reported experiential shifts across the two analysed questions are shown in Table 3. Increased intensity was the most frequently coded shift ($n = 5$), followed by decreased intensity and outsider perspective (both $n = 4$). Insider perspective and no shift were each coded for two participants, and one participant response was coded as descriptive. Pitch, reverb and robot were all three coded across all experiential

shifts (increased & decreased intensity, outsider & insider perspective and deliteralisation). Several participant responses were both coded with an increase and decrease of intensity. One participant's response was both coded with outsider and insider perspective. This suggests that the experience and the perspective of the sentence was dynamic rather than unidirectional.

Increased intensity

> "De modulatie van reverb zorgde bij mij voor het meeste ongemak in mijn beleving [omdat het mijn zin versterkte.]"
> *"The reverb modulation caused me the most discomfort in my experience [because it strengthened my sentence.]"*

Outsider perspective

> "De zin werd meer vervaagt, dit zorgde ervoor dat het leek [alsof ik naar andere personen aan het luisteren was.] Dit kwam voornamelijk voor bij de combinatie pitch & robot."
> *"The sentence became more blurred, which caused it to feel [as if I was listening to other people.] This occurred mainly with the combination of pitch and robot."*

Deliteralisation

> "Pitch, robot en reverb combineren maakte de zin [veel minder een overtuiging die 'echt' kan zijn. Het werd meer een 'geluid' dan een zin, waardoor het bijna geen betekenis meer had.]"
> *"Combining pitch, robot and reverb made the sentence [much less of a conviction that can be 'real'. It became more of a 'sound' than a sentence, causing it to lose almost all meaning.]"*

Metaphor

Spatial metaphors were most frequently coded ($n = 6$), followed by personification ($n = 4$) and atmospheric metaphors ($n = 3$). Weight, sound texture and repetition were each coded for two participants. Across the two questions, reverb was most frequently mentioned in the context of spatial metaphors ($n = 5$). The combination of pitch and reverb was mentioned three times in context of personification.

Spatial

> "Ja, het vervormen van de stem hoger/lager zorgde er voor dat het niet meer mezelf was die ik hoorde, dit gecombineerd met reverb op hoge stand zorgde ervoor dat klonk alsof iemand in [een andere ruimte] dit aan het zeggen was tegen iemand anders."
> *"Yes, distorting the voice higher/lower caused it to no longer sound like myself that I heard, this combined with reverb on a high setting caused it to sound as if someone in [another room] was saying this to someone else."*

Personification

> "Nog meer merkte ik dat mijn eigen stem met reverb het meeste deed, omdat het als een negatief mantra ging voelen [wat een nare priester] tegen mij aan het opdreunen was, alleen was ik het zelf."
> *"What I noticed even more was that my own voice with reverb had the most effect, because it started to feel like a negative mantra [like a sinister priest] reciting at me, only it was me myself."*

Sound texture

> "[De zin voelde op dat moment meer als geluid (een deksel ofzo) in plaats van een gevoel.] Gek genoeg hielp dat."
> *"[At that moment the sentence felt more like a sound (like a lid) rather than a feeling.] Strangely enough, that helped."*

Affective states

Discomfort and humour were most frequently reported ($n = 4$). Pleasantness and seriousness were both reported two times. Followed by hurt, sadness, loneliness and self-compassion ($n = 1$). Across the two questions, pitch was mentioned in the context of humour four times.

Discomfort & loneliness

> "De modulatie van reverb zorgde bij mij voor [het meeste ongemak] in mijn beleving omdat het mijn zin versterkte. [Het gaf een heel eenzaam gevoel.]"
> *"The reverb modulation caused me [the most discomfort] in my experience because it intensified my sentence. [It gave a very lonely feeling.]"*

Humour

> "Begonnen met de pitch aan te passen als eerste van de vijf opties, [dit vond ik vrij humoristisch] wat bij mezelf direct al een stuk luchtigheid creëerde in het geheel."
> *"Starting with adjusting the pitch as the first of the five options, [I found this quite humorous] which immediately created a sense of lightness overall."*

No shift

When asked which modulation produced no shift in their experience of the sentence, panning was most frequently reported ($n = 9$). This is consistent with the one-sample Wilcoxon signed-rank test, which showed no reduction in believability and discomfort, and no increase of acceptance. Robot was reported two times, volume ($n = 1$) and reverb ($n = 1$). Four participants reported that all modulations produced some form of shift. One participant reported that their experience was largely similar across modulations.

DISCUSSION

The present study explored the use of a tangible audio interface as an approach to facilitating cognitive defusion. Four audio modulations were examined for their effect on the impact of a negative self-referential thought. These were measured through believability, discomfort and acceptance. The outcome measures are seen as indicators for cognitive defusion, as the defusion process itself cannot be observed within a single interaction.

Given the small sample size, results should be interpreted with caution. The study is exploratory, therefore the results do not confirm a hypothesis. Results rather show an indication of

a pattern. Interacting with the interface nonetheless lowered both believability and discomfort overall. Acceptance did not change significantly.

Across the four modulations, robot produced the most consistent results with decreases in believability and increases in acceptance. Reverb had an unexpected result; there was an increase in discomfort instead of a decrease. Panning showed no effect. Pitch produced no significant effects, but the change scores varied widely between participants. Volume was not examined in the present study and may still influence the experience in ways not captured here. These results indicate that the modulations did not produce unidirectional effects. The type of audio modulation and individual differences seem to influence the experience of hearing your own negative thought.

Pitch

For pitch no significant effects were found on any of the outcome measures. The qualitative data helps explain this: participants reported both increased and decreased intensity in their experience of the sentence, along with a range of affective states, suggesting the absence of a consistent direction of change. The variety in responses to the pitch modulation could be explained by three factors. First, pitch (also referred to as fundamental frequency, or F0) communicates different emotional qualities depending on its level. Lower pitch has been associated with sadness, while higher pitch has been associated with happiness (among other emotions) [24]. Since participants explored both higher and lower pitch values, some may have found lower pitch more emotionally impactful than other participants. This could be one source of the individual differences observed. A second explanation is that listeners do not all emotionally respond the same way to a given pitch level [24]. Multiple participants reported humour in response to higher pitch, for example, while one participant instead reported seriousness. Third, how representative a recording sounds of someone's own voice varies from person to person [26]. If a difference in baseline affects how participants respond to a pitch modulation, then this could explain why person a is more affected by high pitch than person b. This is supported by the closing interviews; one participant who identified as non-binary found higher pitch more difficult to hear than lower pitch, while another associated lower pitch with age and higher pitch with youth. Personal identity could be one explanation for why pitch modulations are experienced differently.

Robot

Robot showed the most consistent effects across participants, with significant decreases in believability and increases in acceptance. This finding is unexpected, since robot produced the strongest quantitative effects of the four modulations yet drew fewer specific qualitative descriptions than pitch or reverb. Only two participants reported deliteralisation in response to robot [20, 7], an effect that reflects cognitive defusion. With pitch, F0 and the formants shifted together, thereby producing an unnaturally high or low voice that still sounded like a distorted version of the participant's voice. The robot modulation on the other hand altered its timbre. Timbre has been shown to be a primary acoustic cue for vocal emotion perception [40]. If the timbre transformations disrupt the emotional qualities of the voice more effectively, then this could explain why the robot modulation had stronger effects than other modulations. If the voice no longer carries recognisable vocal cues, its function as a negative self-referential statement could therefore weaken. Participants also always included either pitch or robot in their reported combinations, which is consistent with this interpretation, since both modulations alter the vocal qualities most relevant to the emotional impact of the thought. An alternative explanation for the robot modulation could be reduced intelligibility. If participants could not understand the words they had recorded, the sentence may lose its linguistic function as a meaningful statement, which is comparable to deliteralisation. It is unclear whether participants could not understand their recorded thought because of defusion (deliteralisation), distraction, or both. Whether these are distinct processes or overlapping ones cannot be determined from the current data. Distraction in the traditional sense tries to shift the attention away from the emotional stimulus towards something neutral and unrelated. The interface however, is not redirecting the attention to something unrelated, participants stayed focused on the negative thought, at least on an auditory level. The effect could reflect defusion, a shift from linguistic to auditory processing or something in between, this remains an open question and requires further research. What these explanations share is that attention remained directed at the thought rather than away from it. If multiple mechanisms contribute to the observed effect, the interface may reduce the impact of a negative self-referential thought in more than one way.

Reverb

Reverb was expected to create distance between the participant and the thought by simulating a large physical space, which could support the defusion process. Instead, discomfort increased significantly. Believability also increased, although this change was not significant. This is the opposite of the overall pattern, in which the interface lowered both measures. The increase in discomfort could be explained by the longer reverberation time of the used preset. In a recent study longer reverberation times seemed to strengthen emotional characteristics such as 'scary' and 'sad' [36]. In the present study, the qualitative data showed that participants used atmospheric metaphors such as ominous, creepy and haunting to describe reverb. This could be seen as comparable. However, research is not in consensus, another study found that reverberation enhanced feelings of nostalgia, transcendence and wonder. Feelings such as 'unease' were unaffected [52]. Although unease and discomfort are distinct measures, they may be related. Both studies used musical stimuli, not spoken material, and unlike theirs, the recorded material in the present study already carried negative emotional content. Reverberation may therefore have intensified how negative the thought already felt, without transforming it. Reverb tends to make a threatening sound feel more threatening, not more transcendent.

Panning

Panning had no effect on believability, discomfort or acceptance, consistent with the qualitative data in which it was most frequently reported as producing no shift in experience. Panning was implemented using a binaural source panner [13].

This is different from a simple stereo left-right effect. A binaural panner creates movement around the listener's head; however the settings used in this experiment made this effect subtle and barely perceptible through headphones. Binaural audio has been shown to increase emotional impact and engagement in spoken narratives compared to stereo [43]. The implementation used here was relatively simple compared to more advanced binaural or ambisonic techniques, which may explain why no comparable effect was found in this study. Panning also differs from pitch and robot in a more fundamental way. Pitch and robot directly altered the vocal qualities; panning changed only the spatial position of the sound, a property less directly tied to how uncomfortable or believable the thought felt. If panning were extended to a more advanced implementation, it might affect the experience of a negative self-referential thought more strongly.

Overall versus individual

Looking at the overall scores of the interaction we see a decrease in the impact of the thought as a whole even though individual modulations produced varying and sometimes contrasting effects. Several participants showed both increased and decreased intensity within the same interaction, both outsider and insider perspectives, and contrasting affective states. Whether experienced in isolation or in combination, the interface produced a variety of experiences within the same session. There is a downside to this dynamic quality. The increase in discomfort (as seen most clearly with reverb) could carry risks for individuals with higher levels of psychological distress. At this stage of the research, neither the participant nor the researcher can predict which modulation will increase or decrease the impact of the thought. ACT, however, does not aim to reduce distress directly. Instead, the goal is to change the relationship with internal experiences from avoidance to acceptance [20], experiencing uncomfortable thoughts and emotions instead of avoiding them. In that sense, the interface could support this process. One response shows what this could look like:

> 'Vooral de reverb handeling in combinatie met pitch deed voor mij het meeste. Nog meer merkte ik dat mijn eigen stem met reverb het meeste deed, omdat het als een negatief mantra ging voelen wat een nare priester tegen mij aan het opdreunen was, alleen was ik het zelf. Mogelijk voelde ik daardoor wat meer zelfmedelijden dat ik dit over mezelf denk.'
>
> '*Especially the reverb effect combined with pitch affected me the most. I noticed even more that my own voice with reverb had the strongest effect. It started to feel like a negative mantra being chanted at me by an unpleasant priest, except it was myself. That possibly made me feel a bit more self-compassionate for thinking this about myself.*'

Psychological flexibility, the ability to experience uncomfortable emotions without them controlling one's thoughts and actions, is one of the core aims of ACT. Lisboa et al. [31] found that flexibility between immersed and distanced perspectives was associated with adaptive change. In the present experiment, several participants showed shifts consistent with the described flexibility. Their descriptions were coded with contrasting perspectives such as outsider and insider perspective, but also showed both increased and decreased intensity and a variety of affective states. The study by Lisboa et al. researched that flexibility in context of a longer time-span during sessions with a therapist. Here, participants interacted with the interface on their own. Whether the audio interface is able to support this kind of flexibility over the long term, has not been tested and therefore remains an open question.

A recent study on VR-based cognitive defusion in young people with depression and anxiety [5] compared a VR defusion exercise with a non-immersive audio exercise within a single session. Participants in the audio condition listened to a recording of a therapist guiding them through a cognitive defusion exercise. Although participants preferred the VR condition and rated it as more engaging, the audio condition produced more immediate improvements. The authors attributed this outcome to how demanding defusion can be. Individuals have to focus and mentally manipulate a thought. In contrast to the VR environment, the audio condition lacked visual stimuli. This could explain why the audio condition was better equipped to support the process of cognitive defusion. The same rationale can be applied to the interface presented here. Participants engaged directly with recorded sound of their own negative thought. The screen which showed the controls and instructions was kept deliberately plain. Possibly, the attention could therefore be more easily directed to the sound. This is important because distractions could interfere with cognitive defusion. It would be harder to determine whether an observed improvement resulted from being distracted or from a process of cognitive defusion. Future research could compare the audio condition from the VR study against the audio-modulation condition from the present study to explore whether a self-directed exercise based on audio modulation can achieve effects comparable to those of an established guided defusion exercise. However, these two conditions differ at a deeper level. They differ in the content of the exercise itself and whose voice is used, so any different outcome cannot be attributed to guidance alone. It would therefore be more robust if the controlled variable would be the exercise content and the independent variable would be whether the audio is guided or self-directed.

Our findings are consistent with Masuda et al. [34, 33] who found that rapid vocal repetition reduced the believability and discomfort of a negative self-referential thought. In the study done by Masuda et al., the participants repeated the thought aloud themselves; here, they listened to a looped recording of their thought instead. The shared reduction in outcome measures could be ascribed to the same underlying mechanism, namely semantic satiation. Most research on semantic satiation is focused on satiation using active repetition, such as the described study. However, the few studies that do exist show that satiation can also occur through listening [51, 27, 3]. The current study is consistent with the possibility that auditory satiation could also support cognitive defusion, as believability and discomfort of negative self-referential thoughts decreased overall. That said, participants could manipulate the audio of their thought, meaning their engagement was not purely

passive, unlike in the studies described above. Future studies should therefore compare passive and active listening directly.

This could be done by including an unmodulated audio condition, in which participants listen to a recording of their own voice without any audio modulation, allowing passive listening to be tested separately from the modulated condition used here. If we want to compare directly to Masuda et al.'s condition, we would need to match both the content and the pace of rapid vocal repetition. Participants could for instance, listen to a recording of the same rapidly repeated word, instead of repeating it aloud themselves. A further extension could be done by adding an active condition with real-time modulation. Participants would speak the repeated content aloud while being able to manipulate the modulation of their own voice as they hear it through headphones (altered auditory feedback). This would make it possible to test whether combining active production with modulation has an effect beyond the passive, modulated condition used in the present study.

Limitations

One of the limitations of this study is the small sample size ($N = 18$); therefore, results should be interpreted with caution. A sensitivity analysis (G*Power 3.1.9.4) showed that with this sample size, the study had 80% power to detect only large effects ($d_z \geq 0.72$, two-tailed). Smaller effects may therefore have gone undetected, which means that a non-significant result does not necessarily mean there is no effect. Detecting a medium effect ($d_z = 0.5$) would have required about 35 participants, almost twice the current sample. Of the five significant results found among the pre-post and modulation analyses, only one (the believability VAS measure) exceeded the sensitivity threshold of $d_z = 0.72$. The remaining four significant findings had smaller effect sizes than this threshold, and are therefore less certain as a result despite reaching statistical significance. In total 15 statistical tests were run across these analyses without correcting for multiple comparisons (no Bonferroni correction was applied). This increases the risk of Type I errors, meaning one of the five significant results may have occurred by chance. A larger sample size is needed to establish whether these effects hold up.

A related limitation concerns the absence of a baseline exclusion criterion. Masuda et al. [34] excluded participants whose believability and discomfort scores fell below a set threshold, to ensure the thought used was distressing and believable enough to test. This criterion was not applied here, since doing so would have further reduced the small sample size. As a result, some participants may have started with a thought that did not meet that threshold, which could have weakened the study's ability to detect real effects. Applying it in later studies would ensure the thought tested is meaningful for every participant. The acceptance measure raises a separate concern. This measure was included to track whether participants grew more accepting of the thought over the course of the interaction. Believability and discomfort were adapted from Masuda et al. [34], while acceptance was developed in the context of this study and has not been validated elsewhere. The lack of a significant change in acceptance could therefore reflect a genuine absence of effect, or it could reflect a limitation of the measure itself. This could be resolved by validating the measure, or using an established one instead.

This study also relied on a single session with no follow-up assessment. Sessions spread over time, together with follow-up assessments, would show if these effects persist. Repeated sessions have a risk as well. The number of times someone hears a recording of their own voice affects how that voice is recognised [22], which could mean participants grow accustomed to it, reducing the interface's effect over multiple sessions. A single session cannot capture this risk. Future research could also examine whether this effect is different with vocal experience; individuals with extensive vocal training, such as singers, may already relate differently to their own voice.

Participants were free to combine multiple modulations within a single interaction. These combinations were only explored through open-ended questions, not through the same believability, discomfort and acceptance measures used for the individual modulations. It therefore remains unclear whether combining modulations produces something beyond what any single modulation would produce on its own. The open-ended responses were coded through content analysis, as described in the Methods section. This coding was carried out by a single researcher, so no inter-rater reliability could be established. A second coder analysing the same data, or at least a subset of it, would make it possible to check how consistent these codes are.

Implications

Although this study focused on cognitive defusion, the interface's underlying goal is helping people become more flexible and adaptive in how they relate to their emotions and thoughts. The definition of defusion, altering the literal meaning and behaviour-regulatory function of private events without necessarily changing their form, frequency, or situational sensitivity [20, 7], is broad enough to overlap with other therapeutic approaches, and the interface is therefore not limited to ACT exclusively. Distinguishing between defusion and other strategies that share similar goals is not straightforward, as other techniques could possibly facilitate defusion as well. For instance, exposure therapy reduces emotional impact by sustained engagement with a distressing stimulus [15]. This overlaps with the repeated listening to a recording of a negative thought in the present study. It cannot be ruled out that habituation through repeated listening also contributed to the observed effects. It would be valuable to research in what ways defusion overlaps with other strategies and how other strategies outside ACT could support the defusion process.

In practice, the interface could work alongside existing ACT vocal techniques, such as the silly voice exercise [7]. Which technique works best would depend on the client, with a therapist choosing which one works. Participants responded very differently to the individual modulations, which is why that flexibility matters. If no single modulation or vocal technique works the same for everyone, then expanding the range of available tools seems more useful than narrowing it down. The playfulness of the interface may make it easier to sit with

an uncomfortable thought. Exploring and combining modulations can feel more like experimentation than direct confrontation. That playful, experimental quality could make it more accessible for people who find abstract, language-based therapeutic exercises hard to engage with. Metaphor comprehension, for example, has been linked to broader language ability [25]. Even when a metaphor is explained clearly, using it still requires the client to actively imagine a scene or concept. Everyone has different imaginative abilities, hearing the modulated sound of their own voice requires less imaginative effort.

CONCLUSION

This research explored whether modulating a person's own recorded voice could facilitate cognitive defusion and reduce the impact of a negative self-referential thought. Interacting with the interface lowered both the believability and discomfort of the thought overall, although the four modulations affected participants differently. Robot showed the most consistent change, and for reverb, discomfort increased instead of decreased. Panning showed no measurable effect. Pitch produced no significant effect, but the direction of change varied between participants. To the author's knowledge, this is the first study to combine cognitive defusion with this kind of voice modulation. With a small sample size, these findings are a starting point for more research. Future work could clarify which modulations work best for which people, and test the interface against an unmodulated baseline recording. This project draws on several fields at once, from voice perception and how sound shapes emotion to negative self-referential thoughts, ACT, and psychology more broadly.

This thesis is about changing how a thought sounds, which is difficult to convey on paper. For this reason, an interactive version has been created at https://sparrowplague.github.io/distortion-as-acceptance/, where the reader can not only read the research but also listen to each audio modulation and experience how the interface sounds.

REFERENCES

1. Ableton AG. 2025. Ableton Live 12 Suite. https://www.ableton.com. (2025). Version 12.3.2.
2. Auburn Sounds. 2026. Graillon. https://www.auburnsounds.com/products/Graillon.html. (2026). Version 3.1.0.
3. H. L. Baras. 1968. Effect of mode of presentation on semantic satiation. *Psychonomic Science* 11 (1968), 345–346. DOI:http://dx.doi.org/10.3758/BF03328227
4. Y. Barnes-Holmes, S. C. Hayes, D. Barnes-Holmes, and B. Roche. 2001. Relational frame theory: A post-Skinnerian account of human language and cognition. *Advances in Child Development and Behavior* 28 (2001), 101–138. DOI: http://dx.doi.org/10.1016/s0065-2407(02)80063-5
5. I. H. Bell, C. Li, A. Thompson, C. Ellinghaus, S. O'Sullivan, K. A. Reynolds, G. Wadley, Y. Liu, S. Bendall, J. Gleeson, L. Valmaggia, and M. Alvarez-Jimenez. 2025. A virtual reality–based cognitive defusion application for youth depression and anxiety: Mixed methods experimental study. *JMIR Mental Health* 12 (2025), e70160. DOI: http://dx.doi.org/10.2196/70160
6. R. Bennett and J. E. Oliver. 2019. *Acceptance and commitment therapy: 100 key points and techniques*. Routledge.
7. J. T. Blackledge. 2015. *Cognitive defusion in practice: A clinician's guide to assessing, observing, and supporting change in your client*. Context Press.
8. F. W. Bond, S. C. Hayes, R. A. Baer, K. M. Carpenter, N. Guenole, H. K. Orcutt, T. Waltz, and R. D. Zettle. 2011. Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy* 42, 4 (2011), 676–688. DOI: http://dx.doi.org/10.1016/j.beth.2011.03.007
9. T. M. Brinthaupt. 2019. Individual differences in self-talk frequency: Social isolation and cognitive disruption. *Frontiers in Psychology* 10 (2019), 1088. DOI: http://dx.doi.org/10.3389/fpsyg.2019.01088
10. D.-M. Cândea and A. Szentágotai-Tătar. 2020. Cognitive reappraisal as an emotion regulation strategy for shame: Comparing self-distancing to changing self-evaluations. *International Journal of Cognitive Therapy* 13, 1 (2020), 42–53. DOI: http://dx.doi.org/10.1007/s41811-019-00064-4
11. K. Dijkstra, A. Eerland, J. Zijlmans, and L. S. Post. 2014. Embodied cognition, abstract concepts, and the benefits of new technology for implicit body manipulation. *Frontiers in Psychology* 5 (2014), 757. DOI: http://dx.doi.org/10.3389/fpsyg.2014.00757
12. S. Elo and H. Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115. DOI: http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x
13. Envelop. 2025. Envelop for Live. https://github.com/EnvelopSound/EnvelopForLive. (2025). Version 12.2.6. GitHub.
14. L. Fletcher and S. C. Hayes. 2005. Relational frame theory, acceptance and commitment therapy, and a functional analytic definition of mindfulness. *Journal of Rational-Emotive and Cognitive-Behavior Therapy* 23, 4 (2005), 315–336. DOI: http://dx.doi.org/10.1007/s10942-005-0017-7
15. E. B. Foa and M. J. Kozak. 1986. Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin* 99, 1 (1986), 20–35. DOI: http://dx.doi.org/10.1037/0033-2909.99.1.20

16. D. T. Gillanders, H. Bolderston, F. W. Bond, M. Dempster, P. E. Flaxman, L. Campbell, S. Kerr, L. Tansey, P. Noel, C. Ferenbach, S. Masley, L. Roach, J. Lloyd, L. May, S. Clarke, and B. Remington. 2014. The development and initial validation of the cognitive fusion questionnaire. *Behavior Therapy* 45, 1 (2014), 83–101. DOI:http://dx.doi.org/10.1016/j.beth.2013.09.001

17. L. Guo. 2022. Reflect on emotional events from an observer's perspective: A meta-analysis of experimental studies. *Cognition and Emotion* 36, 8 (2022), 1531–1554. DOI:http://dx.doi.org/10.1080/02699931.2022.2134094

18. S. C. Hayes, J. B. Luoma, F. W. Bond, A. Masuda, and J. Lillis. 2006. Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy* 44, 1 (2006), 1–25. DOI: http://dx.doi.org/10.1016/j.brat.2005.06.006

19. S. C. Hayes, J. Pistorello, and M. E. Levin. 2012a. Acceptance and commitment therapy as a unified model of behavior change. *The Counseling Psychologist* 40, 7 (2012), 976–1002. DOI: http://dx.doi.org/10.1177/0011000012460836

20. S. C. Hayes, K. D. Strosahl, and K. G. Wilson. 2012b. *Acceptance and commitment therapy: The process and practice of mindful change* (2nd ed.). Guilford Press.

21. S. C. Hayes, M. Villatte, M. Levin, and M. Hildebrandt. 2011. Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology* 7, 1 (2011), 141–168. DOI:http: //dx.doi.org/10.1146/annurev-clinpsy-032210-104449

22. T. Hosaka, M. Kimura, and Y. Yotsumoto. 2021. Neural representations of own-voice in the human auditory cortex. *Scientific Reports* 11 (2021), 591. DOI: http://dx.doi.org/10.1038/s41598-020-80095-6

23. H. J. Jo, C. Park, E. Lee, J. H. Lee, J. Kim, S. Han, J. Kim, E. J. Kim, E. Kim, and J. J. Kim. 2024. Neural effects of one's own voice on self-talk for emotion regulation. *Brain Sciences* 14, 7 (2024), 637. DOI: http://dx.doi.org/10.3390/brainsci14070637

24. P. N. Juslin and P. Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129, 5 (2003), 770–814. DOI: http://dx.doi.org/10.1037/0033-2909.129.5.770

25. T. Kalandadze, C. Norbury, T. Nærland, and K.-A. B. Næss. 2018. Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism* 22, 2 (2018), 99–117. DOI: http://dx.doi.org/10.1177/1362361316668652

26. M. Kimura and Y. Yotsumoto. 2018. Auditory traits of "own voice". *PLoS ONE* 13, 6 (2018), e0199443. DOI: http://dx.doi.org/10.1371/journal.pone.0199443

27. J. Kounios, S. A. Kotz, and P. J. Holcomb. 2000. On the locus of the semantic satiation effect: Evidence from event-related brain potentials. *Memory & Cognition* 28, 8 (2000), 1366–1377. DOI: http://dx.doi.org/10.3758/bf03211837

28. E. Kross and O. Ayduk. 2011. Making meaning out of negative experiences by self-distancing. *Current Directions in Psychological Science* 20, 3 (2011), 187–191. DOI: http://dx.doi.org/10.1177/0963721411408883

29. G. Lakoff and M. Johnson. 1980. *Metaphors we live by*. University of Chicago Press.

30. J. LeMoult, K. Kircanski, G. Prasad, and I. H. Gotlib. 2017. Negative self-referential processing predicts the recurrence of major depressive episodes. *Clinical Psychological Science* 5, 1 (2017), 174–181. DOI: http://dx.doi.org/10.1177/2167702616654898

31. R. Lisboa, E. Barbosa, I. Moura, J. Salgado, and M. Sousa. 2026. Flexibility between immersion and distancing: Relationship with depressive symptoms and therapeutic alliance. *Psychotherapy Research* 36, 1 (2026), 139–151. DOI: http://dx.doi.org/10.1080/10503307.2025.2451330

32. G. Magni, S. F. M. Pizzoli, E. Scerrati, and C. Repetto. 2025. The impact of embodied and enacted metaphors on cognition and emotions in technology-mediated experiences: A scoping review. *Human Behavior and Emerging Technologies* 2025 (2025), 6032891. DOI: http://dx.doi.org/10.1155/hbe2/6032891

33. A. Masuda, S. C. Hayes, M. P. Twohig, C. Drossel, J. Lillis, and Y. Washio. 2009. A parametric study of cognitive defusion and the believability and discomfort of negative self-relevant thoughts. *Behavior Modification* 33, 2 (2009), 250–262. DOI: http://dx.doi.org/10.1177/0145445508326259

34. A. Masuda, M. P. Twohig, A. R. Stormo, A. B. Feinstein, Y.-Y. Chou, and J. W. Wendell. 2010. The effects of cognitive defusion and thought distraction on emotional discomfort and believability of negative self-referential thoughts. *Journal of Behavior Therapy and Experimental Psychiatry* 41, 1 (2010), 11–17. DOI: http://dx.doi.org/10.1016/j.jbtep.2009.08.006

35. D. S. Mennin and D. M. Fresco. 2013. What, me worry and ruminate about DSM-5 and RDoC? The importance of targeting negative self-referential processing. *Clinical Psychology: Science and Practice* 20, 3 (2013), 258–267. DOI:http://dx.doi.org/10.1111/cpsp.12038

36. R. K. Mo and A. Horner. 2023. The effects of convolution reverberation on the emotional characteristics of musical instrument sounds. *Computer Music Journal* 47, 2 (2023), 57–71. https://muse.jhu.edu/article/944605

37. O. Moran, P. Almada, and L. McHugh. 2018. An investigation into the relationship between the three selves (Self-as-Content, Self-as-Process and Self-as-Context) and mental health in adolescents. *Journal of Contextual Behavioral Science* 7 (2018), 55–62. DOI: http://dx.doi.org/10.1016/j.jcbs.2018.01.002

38. E. M. Murdoch, M. T. Chapman, M. Crane, and D. F. Gucciardi. 2023. The effectiveness of self-distanced versus self-immersed reflections among adults: Systematic review and meta-analysis of experimental studies. *Stress and Health* 39, 2 (2023), 255–271. DOI: http://dx.doi.org/10.1002/smi.3199

39. Novation. Circuit Rhythm. https://novationmusic.com/products/circuit-rhythm. (????). Accessed 22 May 2025.

40. C. Nussbaum, A. Schirmer, and S. R. Schweinberger. 2022. Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience* 17, 12 (2022), 1145–1154. DOI: http://dx.doi.org/10.1093/scan/nsac033

41. P. K. Oleś, T. M. Brinthaupt, R. Dier, and D. Polak. 2020. Types of inner dialogues and functions of self-talk: Comparisons and implications. *Frontiers in Psychology* 11 (2020), 227. DOI: http://dx.doi.org/10.3389/fpsyg.2020.00227

42. R. Rampin and V. Rampin. 2021. Taguette: Open-source qualitative data analysis. *Journal of Open Source Software* 6, 68 (2021), 3522. DOI: http://dx.doi.org/10.21105/joss.03522

43. E. Rodero and I. Rodríguez-de Dios. 2024. The 3D sound power of immersion processing and psychophysiological effects of binaural versus stereo audio stories. *Media Psychology* 27, 5 (2024), 715–736. DOI: http://dx.doi.org/10.1080/15213269.2023.2281545

44. J. Saldaña. 2009. *The coding manual for qualitative researchers*. Sage.

45. L. Smith and R. Klein. 1990. Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16, 5 (1990), 852–861. DOI: http://dx.doi.org/10.1037/0278-7393.16.5.852

46. A. Stefanowska and S. K. Zieliński. 2024. Spatial sound and emotions: A literature survey on the relationship between spatially rendered audio and listeners' affective responses. *International Journal of Electronics and Telecommunications* 70, 2 (2024), 293–300. DOI: http://dx.doi.org/10.24425/ijet.2024.149544

47. J. A. Stoddard and N. Afari. 2014. *The big book of ACT metaphors: A practitioner's guide to experiential exercises and metaphors in acceptance and commitment therapy*. New Harbinger Publications.

48. A. M. Tackman, D. A. Sbarra, A. L. Carey, M. B. Donnellan, A. B. Horn, N. S. Holtzman, T. S. Edwards, J. W. Pennebaker, and M. R. Mehl. 2019. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of Personality and Social Psychology* 116, 5 (2019), 817–834. DOI: http://dx.doi.org/10.1037/pspp0000187

49. E. B. Titchener. 1910. *A textbook of psychology*. Scholars' Facsimiles & Reprints.

50. M. P. Twohig, M. E. Levin, and J. M. Petersen (Eds.). 2023. *The Oxford handbook of acceptance and commitment therapy*. Oxford University Press. DOI:http://dx.doi.org/10.1093/oxfordhb/9780197550076.001.0001

51. M. S. Vitevitch, J. W. Ng, E. Hatley, and N. Castro. 2021. Phonological but not semantic influences on the speech-to-song illusion. *Quarterly Journal of Experimental Psychology* 74, 4 (2021), 585–597. DOI: http://dx.doi.org/10.1177/1747021820969144

52. H. Wilkie and P. Harrison. 2025. Reverberation time and musical emotion in recorded music listening. (2025). DOI:http://dx.doi.org/10.17863/CAM.110863

APPENDIX

SCRIPT EXPERIMENT

Instruction Text

English

Welcome. In this session, we are examining how people experience a short, self-recorded negative self-referential thought while interacting with an audio interface. By a negative self-referential thought, we mean a brief first-person statement about yourself (3–7 words) that currently feels believable and may evoke some discomfort.

You will treat your sentence as a sound/object whose properties you can adjust. You formulate one sentence, record it (up to three attempts), and then modify the sound using four controls. We will first practise briefly with the device so it is clear what you need to do.

Before and after the task, you will complete a few short questions about your experience (believability, distress/discomfort, and acceptance), and a few open questions.

The content of your sentence remains private; only you will hear the recording. The audio is deleted at the end of the session. You may pause or stop at any time.

If you have any questions, please feel free to ask before we begin.

Dutch

Welkom. In deze sessie onderzoeken we hoe mensen een korte, zelf ingesproken negatieve zelfreferentiële gedachte ervaren tijdens interactie met een audio-interface. Met een negatieve zelfreferentiële gedachte bedoelen we een korte eerste-persoons uitspraak over jezelf (3–7 woorden) die je nu geloofwaardig vindt en die enig ongemak kan oproepen.

Je behandelt je zin als een geluid/object waarvan je eigenschappen kunt moduleren. Je formuleert één zin, neemt die op (maximaal drie pogingen) en past daarna het geluid aan met vier knoppen. Vooraf oefenen we kort met het apparaat zodat duidelijk is wat je moet doen.

Voor en na de taak vul je enkele korte vragen in over je ervaring (geloofwaardigheid, spanning/ongemak en acceptatie) en over de duidelijkheid en bruikbaarheid van de interface.

De inhoud van je zin blijft privé; alleen jij hoort de opname. De audio wordt aan het einde van de sessie verwijderd. Je kunt op elk moment pauzeren of stoppen.

Heb je vragen, stel ze gerust voordat we beginnen.

INSTRUCTION THOUGHT SELECTION

Standardised Instructions

Dutch

Lees voor:

Formuleer nu één korte negatieve ik-uitspraak (3–7 woorden, tegenwoordige tijd, één idee) die je nu geloofwaardig vindt en minstens enigszins belastend is. De inhoud blijft privé.

Geloofwaardigheid betekend dat je de negatieve ik uitspraak als waar beschouwd, als iets wat echt waar is. Iets wat je gelooft over jezelf.

Voor de onderzoeker: Als participanten vragen of het goed, is of hulp willen bij hun negatieve gedachte. Vraag dan niet naar de inhoud, maar vraag of het geloofwaardig is en belastend is. Het is iets wat je kan belemmeren in het dagelijks leven of ongemakkelijk voelt.

English

Read out loud:

Now formulate one short negative self-statement (3–7 words, present tense, one idea) that you currently find believable and at least somewhat distressing. The content remains private.

Believability means that you consider the negative self-statement to be true, something you genuinely believe about yourself.

For the researcher: If participants ask whether it is “okay” or want help, do not ask about the content. Instead, ask whether it is believable and distressing. It should be something that could interfere with daily life or feel uncomfortable.

Thought Selection Worksheet

*Please complete this worksheet before the audio interaction. Your sentence content stays **private**. You do not have to tell it to the researcher.*

1) What counts as a suitable thought?

A negative self-referential thought is:

- About you (uses I / me / my).
- A brief belief/appraisal (not just a feeling).
- Present or near-future (e.g., I am… / I can't… / I will…).
- One idea, 3–7 words, speakable in 1–2 seconds.
- Believable to you right now and at least distressing.
- Avoid names or identifying details (e.g., real names, initials, addresses, workplaces). Use generic words instead (e.g., "someone", "work", "home").

2) Build your sentence (use one template or write your own)

Domain	Template
Identity / worth	I am [negative trait]
Ability / competence	I can't [verb]
Acceptance/ likeability	I am [unlikeable / not welcome / don't belong]
Deservingness	I don't deserve [X]
Responsibility / blame	I [verb] things / I make [noun] worse
Prediction	I will [negative outcome]
Control / agency	I have no [noun] / I can't [verb] this
Appearance / body	I look [negative]
Morality / character	I am [moral judgement]

Draft your own:

_____

3) Final sentence (write clearly)

_____

Werkblad Gedachte-selectie

Vul dit werkblad in vóór de audio-interactie. De inhoud van je zin blijft privé; je hoeft deze niet met de onderzoeker te delen.

1) Wat is een geschikte gedachte?

Een negatieve zelf-referentiële gedachte is:

- Over jouzelf (gebruikt ik / mij / mijn; bij voorkeur met ik als onderwerp, bv. “Ik ben…”.
- Een korte gedachte/oordeel (niet alleen een gevoel).
- In de tegenwoordige tijd of nabije toekomst (bv. Ik ben… / Ik kan niet… / Ik zal…).
- Eén idee, 3–7 woorden.
- Geloofwaardig voor jou op dit moment én minstens enigszins belastend/spannend.
- Vermijd namen of identificerende details (bijv. echte namen, initialen, adressen, werkplekken). Gebruik in plaats daarvan algemene termen (bijv. ‘iemand’, ‘werk’, ‘thuis’).

2) Bouw je zin (gebruik een template of maak je eigen zin)

Domain	Template
Identiteit / waarde	Ik ben [negatieve eigenschap]
Vermogen / competentie	Ik kan niet [werkwoord]
Acceptatie	Ik ben [onbemind / niet welkom]
Verdienen/ waard zijn	Ik verdien geen [X]
Verantwoordelijkheid / schuld	Ik [werkwoord] dingen/
Voorspelling	Ik zal [negatieve uitkomst]
Controle / handelingsvermogen	Ik heb geen [zelfst. nw.]
Uiterlijk / lichaam	Ik zie er [negatief oordeel]
Moraliteit / karakter	Ik ben [morel oordeel]

Schets je eigen zin:

_____

3) Definitieve zin (schrijf het duidelijk op)

_____

CODEBOOK CONTENT ANALYSIS

This codebook was developed to ensure transparency and replicability of the qualitative analysis. It contains definitions, inclusion criteria and examples for each category, and is organised into three sections reflecting the three analyses conducted: experiential shifts, metaphor and affective states. Codes were applied to the data using Taguette [42].

General coding rules

- A single fragment may be coded under multiple categories across the three analyses (experiential shifts, metaphor & affective responses) simultaneously.
- Highly conventionalised spatial terms for pitch such as 'high' and 'low' were not coded as spatial metaphors, as these terms are directly adopted from the interface and do not reflect spontaneous metaphorical language use by the participant.

Experiential shifts

This analysis encompasses shifts in perspective, intensity and perceived realness of the thought as reported by participants during interaction with the interface.

Increased intensity categorises phrases and words that display an increase of intensity experienced during the experiment. The phrases that are used to describe these are descriptive. Affective states are excluded from this categorisation and will be coded separately. This way we can code whether or not there was an increase or decrease of intensity during the experience, but in parallel also code which affective states are present. To illustrate this, an example response will be coded that includes an affective state and an increase in intensity. Only the section that describes the increased intensity is coded between *[brackets]*.

For example: 'De modulatie van reverb zorgde bij mij voor het meeste ongemak in mijn beleving *[omdat het mijn zin versterkte.]* Het gaf een heel eenzaam gevoel.' (The reverb modulation caused me the most discomfort in my experience *[because it strengthened my sentence.]* It gave a very lonely feeling.).

Decreased intensity categorises phrases and words that display a decrease of intensity experienced during the experiment. The phrases that are used to describe these are descriptive. Affective states are excluded from this categorisation and will be coded separately. This way we can code whether or not there was an increase or decrease of intensity during the experience, but in parallel also code which affective states are present. To illustrate this, an example response will be shown that includes a decrease in intensity.

For example: 'Het combineren van 2 modulaties zorgde ervoor dat mijn beleving afnam.' (combining two modulations caused my experience to diminish)

Outsider and insider perspective refer to the two viewpoints from which participants may experience the thought. Perspective shifts are considered relevant to cognitive defusion, as the ability to observe thoughts from an external viewpoint is a core component of defusion within the ACT framework [20, 7]. Both perspectives are coded because research suggests that the dynamic alternation between them, rather than a linear shift from one to the other, is associated with greater flexibility and adaptive change [31]. Coding both perspectives allows for an assessment of whether the modulations produced a shift in perspective, and in which direction.

Outsider perspective is categorised as adopting the position of an observer rather than the author of the thought. Within the ACT framework, this is described as stepping back from the meaning of verbal processes and witnessing them from the point of view of an observer, consistent with the concept of self-as-context [20, 7]. Self-distancing research additionally documents this shift, showing that observing oneself from an external viewpoint reduces the intensity and duration of emotional responses [17, 28]. Drawing on both frameworks allows for a broader operationalisation of the category, capturing any instance in which the participant adopted an external perspective toward the thought, regardless of the underlying mechanism.

For example: 'het leek alsof ik naar andere personen aan het luisteren was' (it felt as if I was listening to other people); 'niet meer mezelf die ik hoorde' (no longer myself that I heard).

Insider perspective is categorised as the thought was experienced as one's own, from a first-person viewpoint. Self-immersion research, describes this as experiencing events through one's own eyes, with an absence of psychological distance [38]. Although ACT describes cognitive fusion as treating thoughts as literal truths rather than mental events, it does not explicitly operationalise the perspective from which this occurs and therefore cannot be coded as such. This category is nonetheless included for the reasons outlined in the introduction to this section. Expressions describing the location of the thought, such as 'in my head', are excluded because they describe where the thought is perceived (they are coded as spatial) rather than the perspective from which it is experienced.

For example: 'alleen was ik het zelf' (only it was me myself); 'it felt like I actually said it…'.

Deliteralisation refers to instances where the sentence felt less meaningful or less real. This is consistent with ACT's description of deliteralisation as a temporary loss or weakening of word meaning as a core component of cognitive defusion [20, 7].

For example: 'het werd meer een geluid dan een zin, waardoor het bijna geen betekenis meer had' (it became more of a sound than a sentence, causing it to lose almost all meaning); 'it did not feel real'.

Metaphor

The metaphor analysis assesses the language participants spontaneously used to describe their experience of the experiment. A metaphor is understood here as an expression in which language from one domain is used to describe an experience in another [29]. Some terms are theoretically metaphorical in origin (such as 'high/low' for pitch), but as participants adopted these directly from the interface rather than generating them independently, they were not coded. The following categories were identified inductively from the data.

Spatial is categorised as metaphors of location or distance, describing where the thought was perceived to be. This includes descriptions of the thought as positioned in a specific place or moving closer or further away. Coded when the fragment explicitly describes location or distance, not when it describes a personification, or sound texture. To illustrate this, an example response will be coded that includes a personification and a spatial metaphor. Only the spatial metaphor is coded between *[brackets]*.

For example: 'dit gecombineerd met reverb op hoge stand zorgde ervoor dat klonk alsof iemand *[in een andere ruimte]* dit aan het zeggen was tegen iemand anders' (this combined with reverb on a high setting caused it to sound as if someone *[in another room]* was saying this to someone else).

Personification is a metaphor where the participant describes the modulated audio as though it were spoken or produced by a person, entity or separate agent with its own character, voice or intention. This differs from outsider perspective, which describes a shift in the position of the participant as observer. For example: 'dit zorgde ervoor dat het leek alsof ik naar andere personen aan het luisteren was' (This caused it to feel as if I was listening to other people). Personification instead refers to a shift in how the modulated audio itself is characterised namely, as having a separate identity regardless of the participant's own perspective.

For example: 'een nare priester die tegen mij aan het opdreunen was' (a sinister priest reciting at me); 'een stem in mijn hoofd die tegen mij praatte' (a voice in my head talking to me).

Weight is a metaphor of physical weight (heavy or light) applied to the experience of the sentence. The intensity categories however, describe an increase or decrease in how strongly the sentence was experienced.

For example: 'omdat het mijn zin versterkte' (because it intensified my sentence); 'made me forget or alleviate the thought'. Weight on the other hand refers specifically to the use of physical weight as a metaphor. For example: 'het gaf gewicht aan de zin' (it gave weight to the sentence).

Atmospheric are instances where the modulations evoked a particular mood or atmosphere.

For example: 'made the thought feel more ominous'; 'surreal yet more haunting'.

Sound texture metaphors describe the perceived quality of the modulations.

For example: 'de zin werd meer vervaagd' (the sentence became more blurred); 'de zin voelde meer als geluid, een deksel ofzo' (the sentence felt more like a sound, like a lid).

Repetition are metaphors in which the looping of the audio is described as a recurring or cyclical experience.

For example: 'omdat het als een negatief *[mantra]* ging voelen' (because it started to feel like a negative mantra); 'made the sentence feel like a *[repeating mantra]* during a dream or nightmare'.

Affective states

Affective states responses were coded based on emotional states explicitly reported by participants, as well as affective states inferred by the researcher from the language used [44]. Below is the list of all affective states that were coded.

Humour defined as words and/or phrases expressing amusement or finding the modulation funny.

For example: 'ik moest lachen' (I had to laugh); 'dit vond ik vrij humoristisch' (I found this quite humorous).

Discomfort defined as words and/or phrases expressing a general sense of unease or discomfort in response to the modulation.

For example: 'meer confronterend en beladen gevoel, ongemak' (more confronting and heavy feeling, discomfort).

Hurt defined as words and/or phrases expressing a sense of pain or being hurt in response to the modulation.

For example: 'more hurtful'.

Loneliness defined as words and/or phrases expressing a sense of loneliness or isolation in response to the modulation.

For example: 'het gaf een heel eenzaam gevoel' (it gave a very lonely feeling).

Pleasantness defined as words and/or phrases expressing a general sense of agreeableness or positive feeling in response to the modulation.

For example: 'veel prettiger' (much more pleasant); 'het positiever werd' (it became more positive).

Sadness defined as words and/or phrases expressing a sense of sadness in response to the modulation.

For example: 'during a moment of sadness'.

Seriousness defined as words and/or phrases expressing that the modulation made the sentence feel more serious.

For example: 'it seemed more serious'.

Self-compassion defined as words and/or phrases expressing compassion or sympathy toward oneself in response to the modulation.

For example: 'mogelijk voelde ik daardoor wat meer zelfmedelijden dat ik dit over mezelf denk' (possibly I felt more self-compassion that I think this about myself).

Table 4: Content analysis per participant: modulations mentioned, experiential shifts, metaphors and affective states.

P	Modulations	Experiential shifts	Metaphors	Affective states
P1	Pitch, Reverb, Panning, Volume	Decreased intensity	Spatial	Pleasantness
P2	Reverb	Increased intensity, Decreased intensity	–	Discomfort, Loneliness
P3	Pitch, Robot, Reverb	Increased intensity, Decreased intensity	–	–
P4	Pitch, Reverb, Volume	Increased intensity, Descriptive	Personification	–
P5	Pitch	No shift	–	Humour, Pleasantness
P6	Pitch, Robot, Reverb, Panning	Deliteralisation	Spatial, Personification	Discomfort
P7	Pitch, Reverb	Increased intensity	–	–
P8	Pitch, Robot, Reverb	Outsider perspective	Spatial, Sound texture	–
P9	Pitch, Reverb	–	–	Humour
P10	Pitch, Robot, Reverb	–	Sound texture	Humour
P11	Pitch, Reverb	Outsider perspective	Spatial, Personification	Humour
P12	Pitch, Reverb	Outsider perspective, Insider perspective	Personification, Repetition	Discomfort, Self-compassion
P13	Pitch, Reverb	–	Weight, Atmospheric	Discomfort
P14	Robot, Reverb	Decreased intensity	Spatial, Atmospheric	–
P15	Pitch, Robot, Reverb	Insider perspective	Weight, Atmospheric, Repetition	Sadness, Seriousness
P16	Pitch	Outsider perspective, No shift	–	–
P17	Robot, Reverb	Deliteralisation	Spatial	–
P18	Pitch, Robot	Increased intensity	–	Hurt, Seriousness