DOI: 10.61782/fa.2023.0524

forum acusticum 2023

EXPLORING REVERSE CORRELATION TO HACK THE MENTAL REPRESENTATION OF JOY IN SPEECH SIGNALS

Julian Moreira[1*] **Rozenn Nicol**[1] **Laetitia Gros**[1]
Nicolas Voisine[1]
[1] Orange Labs, Lannion, France

ABSTRACT

Reverse correlation is a psychophysical method that aims at extracting the salient features of stimuli that drive a judgement, leading to the mental representation (what is referred to as the classification image) of this judgment. It can be seen as a mapping of a low-level parameter space (i.e. stimuli) to a high-level feature space (i.e. judgment), on the basis of randomly generated stimuli. In this paper, we focus on the mental representation of joy in audio speech, considering three acoustic parameters: pitch, duration and loudness. Particularly, we investigate the influence of the reverse correlation method variables (choice of low-level parameters, base stimulus, method of random variation to generate stimuli, number of trials) on the resulting classification image. A first experiment reproduces the protocol of a literature study. The results raise several questions, leading to a second experiment with a modified protocol, seeking to narrow the stimulus space while refining the search. Noticeable differences between the classification images obtained in the two experiments are pointed out. Possible improvements to the reverse correlation method (in terms of efficiency and reliability) are outlined and discussed.

Keywords: *reverse correlation, social perception, speech, two-alternative forced choice, psychophysics*

**Corresponding author*: julian.moreira@orange.com.

1. INTRODUCTION

The world appears to our senses as a huge accumulation of continuous, entangled signals. Despite this, in daily life, we constantly and rapidly derive high-level information from our environment, making us able to interact with it. This process of inference can be roughly modeled as a mapping process, between, on one side, low-level properties of the input signals and, on the other side, high-level representations we have in mind. For instance, if the input is a speech signal, we may try (consciously or not) to match its acoustic properties with internal representations to label it with an emotion. Having access to this mapping, and the associated mental representation, would open to many advantages, helping not only to understand better, but also to model and synthesize human perception, behavior and interactions.

Reverse correlation is a psychophysical method that aims at extracting such mental representations. More precisely, it seeks to estimate a mental representation by building a so-called *Classification Image or CI*. The notion of reverse correlation aroused in opposition to regular correlation (”experimenter-centred” approach), where stimuli with controlled properties are presented to subjects. Responses are then collected and an average response is derived for each specific property, allowing to model the response as a function of the stimulus properties. With reverse correlation (”subject-centred” approach), the opposite is done. Stimuli are generated with random properties, i.e., with no preconceived assumptions. For each trial and each subject, a different stimulus is presented. What is looked in the subject’s responses is a ”sign that something particular happened” [1]. All the stimuli which are associated to a specific response are then averaged, leading to extract the salient features which trigger this reaction.

Asserting that a given response conveys a specific mental representation, the mean stimulus corresponds to what we call a CI.

The benefits of reverse correlation are manifold: first, it allows from the individual point of view to analyze and better understand the link between low-level properties of our direct environment and high-level perception features (e.g., visual perception of lightness [2], of face emotions, gender or individual facial traits [3]). Second, from a more collective point of view, cultural attributes can be highlighted by agglomerating individual CIs, as well as intercultural differences (for instance, concerning the perception of face emotions [4]). Third, perception discrepancies between healthy individuals and mentally disordered or pathological individuals (e.g., speech and music perception with autistic vs. non-autistic individuals [5]) may be studied. Finally, to enhance computer-to-human or distant human-to-human interactions, it is possible to make one's signal (human or machine) consistent with the CI extracted from the other one, as suggested in [6], where authors set up a reusable audio filter as CI, embedding acoustic properties of honesty or certainty in speech.

In this paper, we assess the process of building a CI through the case study of mental representation of joy in audio speech. As reverse correlation relies on a specific experimental paradigm, we examine to what extent tuning method parameters influence the quality of the resulting CI. Starting from recent publications about reverse correlation applied to speech signals [6–8], we set up an exploratory study through two iteratively designed experiments: the first one roughly reproduces a state-of-the-art protocol, the other one introduces changes, relatively to the stimuli generation and the CI construction. After an insight into the reverse correlation method in Section 2, we present our experiments in Section 3 and 4, leading to a global discussion on reverse correlation and possible future works in Section 5. We conclude in Section 6.

2. FUNDAMENTALS OF REVERSE CORRELATION

2.1 General philosophy

Reverse correlation is a data-driven method. Participants are presented a set of randomly varying stimuli, for which the experimenter has no a priori hypothese on participant's responses. The goal being to identify those properties of the stimuli that will determine a given judgment (for instance facial traits that will lead an outside observer to judge a face as joyful), the fundamental idea of reverse correlation is to generate a set of stimuli (in the chosen example, pictures of human faces), which aim at representing in the most exhaustive way the diversity of facial expressions, without being limited to the expression of joy. Thus, participants are asked to select the faces which convey joy, following, for example, a two-alternative forced choice protocol (2AFC) [9]. This strategy is inspired from signal detection theory [9], but it is noteworthy that reverse correlation focuses on the case of false alarms, i.e. when the participant detects a signal whereas the stimulus contains only noise. More exactly, it is intended to explain which features of the noise lead to erroneous detection of a signal in a pure noise. By averaging all the stimuli selected as joyful, the salient features driving the perception of joy are highlighted. This average pattern defines a CI associated to the judgment of interest. It may be computed either for one individual, or for a group of individuals.

When looking for the features governing a judgment, it is difficult to have a priori a clear overview of all the potential features. Most of the time, the field of possibilities is very wide. Moreover, unexpected feature may contribute to the judgment. One strength of reverse correlation relies on the randomization of stimuli. The sole assumption that is made concerns the properties of the random distribution from which stimuli are generated, namely the type of distribution (e.g. gaussian, uniform) and the range of variation (e.g. the mean and the standard deviation in the case of a gaussian generation). A second strength relates to the fact that, when participants are making their judgment during the experiment, they are free to use whatever criteria, in full spontaneity [9]. Most often, they may not even be aware of their criteria.

The reverse correlation paradigm unfolds in 4 main steps, detailed below: 1) production of random stimuli; 2) collection of participant's judgments; 3) computing of CIs from participant's reponses; 4) evaluation of CIs.

2.2 Stimuli

Stimuli are obtained by applying random variations to a set of base stimuli, that are designed to be as neutral as possible. At least they represent a kind of mean anchor from which variations can spread. In the case of faces, base stimuli are average faces taken from databases [3, 4]. As for studies on speech prosody, base stimuli may be obtained by flattening speech samples (e.g., words or pseudo-words) [6]. Then, random variation of base stimuli may be achieved in various ways, either by superim-

posing a random noise (e.g. white noise, sine-wave noise or Gabor noise [9]) on them, or by random data generation following a predefined model of stimulus synthesis [6]. All these parameters of variation determine the level of exhaustiveness with which stimuli illustrate the phenomenon under evaluation. However, the wider the stimulus space is, the more trials are needed to explore it. As the most extreme example, a study investigated the mental representation of the letter "S" with stimuli only containing white noise (i.e., without any base stimulus) [10]. The required number of trials was 20,000.

2.3 Experimental paradigm

During the experiment, the participant is asked to judge the stimuli according to the dimension under interest. For each trial, either one single stimulus or a set of stimuli is presented. The task is designed accordingly. For instance, in a 4AFC (four-alternative forced choice) paradigm, the participant is presented one stimulus and he(she) has to rate it on a 4-point scale. In a 2IFC (two-image forced choice) paradigm, a pair of stimuli is presented, one stimulus being obtained with the mathematical inverse of the noise pattern used for the other stimulus. For each pair, the participant has to select the stimulus that matches the best his(her) mental representation.

2.4 Deriving CIs

The goal of reverse correlation is to extract the salient features which drive a given judgment. These features are brought to light by averaging the stimuli which are selected by participants in accordance with the dimension under assessment. If stimuli are obtained by superimposing noise on a base stimulus, noise patterns are averaged. If stimuli are synthesized by a parametric model, in which the randomization is limited to the model parameters, averaging is performed on the parameter patterns themselves. The resulting CI is considered as a proxy for the expected mental representation. The CI may be computed for each individual or for a group of individuals.

2.5 Evaluating CIs

The reliability with which CIs estimate the mental filter of interest may be assessed by conducting a second experiment, where CIs are used to generate stimuli and participants evaluate a new dimension related to the one evaluated in the original experiment (e.g., investigating the trustworthiness in a male face from CIs derived for male and female faces). Statistical analysis of the information conveyed by CIs can also provide useful insight [11].

3. EXPERIMENT 1

3.1 Presentation

Our first experiment is inspired by [6], where Goupil et al. extract perception of honesty and certainty in speech. We aim to capture the representation of joy in speech with a similar method, and to assess the reliability of the resulting perceptual filter. Similarly to Goupil et al., we generate stimuli variations along three acoustic parameters: pitch, duration and loudness. These parameters are usually considered to be of primary importance in prosodic analysis of emotions in speech (e.g., [12, 13]).

3.2 Method

3.2.1 Participants

Forty native French listeners participate in the study (mean age = 41.9 +/-11.8, 20 females). All report a normal audition. Participants are external to the laboratory, recruited from a database they deliberately registered on (giving their consent to take part in the experiment), and receive a 20 EUR voucher to compensate for their time.

3.2.2 Stimuli

First, twenty monophonic reference stimuli are produced with two speakers (1 male and 1 female), each uttering 10 bi-syllabic pseudo-words: *bazin, bivan, bodou, dadon, dejon, dobue, gibue, vagio, vevon, vizou*. Using pseudo-words ensures not to embed emotional content in semantics. The selected pseudo-words are designed to be phonetically representative of French language, in terms of diversity and frequency of the syllables [6]. Second, a flattening phase is performed on the reference stimuli, along pitch, duration and loudness. For each parameter, the flattening operation results in a constant stimulus, not only all the way along its course, but also accordingly with the other reference stimuli. Flattening is performed by CLEESE, a voice transformation toolbox designed for reverse correlation, and implementing a phase-vocoder digital audio technique [14], used for pitch shifting and time stretching. The processing is based on short-term Fourier Transform (STFT), operating on 20ms-frames. Duration is flattened per syllable, while pitch and loudness are flattened every 20ms. Third, from each flat stimulus, and for each subject, 80 random variations are generated, also

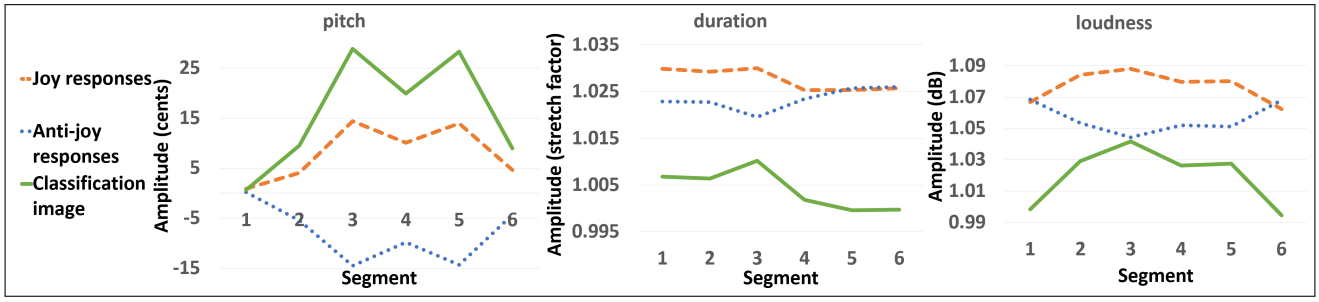

Figure 1. Average CIs of joy in speech in experiment 1, across all subjects, along three parameters: pitch (left), duration (middle) and loudness (right). The medium-dashed line is the mean pitch/stretch factor/loudness of the stimuli classified as joyful, the small-dashed line is the mean pitch/stretch factor/loudness of the discarded stimuli (so-called "anti-joy"), and the solid line is the resulting CI for the considered parameter.

with CLEESE, along the 3 acoustic parameters. Each stimulus is divided into 6 segments of equal length, and each segment is assigned with a random pitch, duration and loudness variation value, resulting in a stimulus with dynamic profiles for the 3 parameters. In this experiment, randomness is ruled by a normal distribution for pitch (SD = 100 cents, clipped at +/-2.2 SD) and loudness (SD = 1.7 dB, clipped at 2.2 + /- SD). For duration, we point out that time stretching is not a symmetrical operation around the neutral value. E.g., time stretching by a factor 0.5 divides the duration of the stimulus by 2; but if on the contrary we want to multiply the duration by 2, the time stretching factor is 2, and not 1.5. Therefore, we use a uniform distribution (ranging from 0.8 to 1.25 stretching factor). Still following [6], we choose these values to produce stimuli as natural as possible. Subjects are divided into two groups of 20, one is assigned to the male speaker stimuli, the other to the female speaker stimuli. All in all, a total of 80 random variations x 10 pseudo-words x 2 speakers x 20 subjects = 32 000 stimuli is generated.

3.2.3 Apparatus

The experiment is set up on a laptop Dell Alienware 15 R4 (Intel Core i7, 8 GB of RAM, Windows 10), equipped with a soundcard Focusrite Scarlett 6i6. Sound is rendered by Sennheiser HD 650 open headphones.

3.2.4 Procedure

Stimuli are presented in pairs, following the two-alternative forced choice protocol (2AFC) [9], and can be listened to only once. For each pair, subject is asked to choose the one that he(she) perceives to be the most joyful. After a quick training phase of 4 pairs, subject is presented a total of 440 pairs: 10 blocks of 40 pairs (i.e., 80 stimuli), one block for each pseudo-word, plus an additional repetition block, in order to measure subject's self-consistency. The order of the 10 blocks is randomly shuffled for each subject. The choice of the repetition block varies from one subject to another, so that each pseudo-word is repeated twice per speaker. After each block, the subject is invited to take a break. The whole experiment takes about 1h.

3.3 Results

The CI consists in three different dynamic profile curves: one for the pitch, one for the duration and one for the loudness (see Fig. 1). In the case of the stretch factor and the loudness, the CI is computed by dividing the mean of the stimuli classified as joyful by the mean of the discarded stimuli. As for the pitch, the CI is obtained by subtracting the mean of the discarded stimuli from the mean of the stimuli classified as joyful. The mean is computed after removing the repetition block. It is calculated across all subjects, meaning that a shared image of joy in speech is obtained . On the pitch curve, starting from 0, we observe a global increase (with two peaks at segments 3 and 5), going back down on the last segment. The highest pitch increase reaches 28.8 cents, i.e., a little bit less than a sixth of tone. For the duration, we observe at start a progressively decelerating voice (stretch factor greater than 1), until a peak of slowness at segment 3, then a sudden acceleration from segment 3 to 4 (i.e. stretch factor lower than 1), moderately progressing until the end. Despite these variation patterns, we note that all the values are close to 1. Concerning the loudness, we observe a profile similar to the pitch, i.e. an increase in the middle segment, with

the starting and ending segments around the neutral value 1. However, again, the variations are quite small.
It should be noticed that these CI curves constitute a prosodic profile of joy for each parameter, and could be reapplied as a 'prosodic filter' on new audio samples. This is the classic procedure in reverse correlation experiments to further study CIs, and assess their consistency with mental representations of the subjects (as explained in Section 2). As, in this paper, we strictly focus on the building phase of the CI, we postpone the validation step to a later study. We solely conduct an informal perceptual assessment session between authors, by applying the CI to the flattened and reference stimuli.
As for the repetition block, we compare the answers to the ones obtained with the original block for each subject, and calculate the percentage of identical responses. Across all subjects, the average self-consistency is 68.2% +/-11, which corresponds to the value found in [6].

3.4 Discussion

As mentioned in the previous section, variations in CI are globally small. In the listening test, none of the authors were able to perceive a change between the base stimuli and the modified ones. A future experiment will investigate this issue. Anyway, this observation can be explained in several ways. As mentioned in [9], "the numerical values of the average patterns are typically very small (due to averaging patterns that contain little signal and mostly random values)". Removing the mean of the discarded answers to the mean of the selected ones is one way to reduce this effect, but it may be insufficient. Furthermore, for pitch and loudness, the choice of a normal distribution implies that most of the generated variations are close to the neutral value. This is even reinforced by the deliberately chosen small range of variations (for the three parameters), in order to keep the stimuli as natural as possible. However, the will to naturalness 'at any cost' may not suit the joy emotion. Large variations of the three parameters may produce a lot of absurd stimuli, but sometimes, a combination of a small variation of one parameter, with a large variation in another, could lead to a better stimulus candidate. An informal test with a uniform random distribution for all the three parameters and a larger range of variations supported this hypothesis. For pitch and loudness, we get profile curves similar to Fig. 1, but with higher peaks. For duration, the stretch factor remained near 0.
Another possible explanation is the flattening process at the stimuli generation. In our experiment, this process led to audible artefacts (e.g., robotic voices, as similarly reported in [6] by a few subjects – 3 out of 115 participants), possibly resulting in an even more challenging assessment task for the subjects. Furthermore, no clear justification of this flattening process is provided in [6], nor in other publications invoking the same requirement of "flat values" [7,8]. We initially assumed that it was a way to emotionally neutralize the stimuli, as an analogous process to what is done in reverse correlation experiments investigating emotions in faces, where base stimuli is obtained by averaging pictures taken from a database (see Section 2) [3,4]. But the process of computing an average of real stimuli is not equivalent to the flattening process described in [6], where the prosodic profile, artificially tuned to a constant value, tends to move the stimulus away from real stimuli. Thus, its relevance may be questioned.
Another question of importance is the number of trials. In [9], a typical number of 300-1000 trials is reported for noise-based reverse correlation related to visual perception of faces. In acoustics, to our knowledge, no study has been conducted to determine the optimal number of trials. In our experiment, to check whether 400 trials are enough to reliably estimate a CI, we introduce a new metric aiming at measuring the distance between the CI obtained at the end of the experiment (i.e. the "global" CI) and a CI would have been obtained with less trials (i.e. a "partial" CI). We call this metric "CI stability", as it evaluates the extent to which partial CIs are close to the final CI, as a function of the number of trials. For instance, to evaluate the CI stability for 100 trials, we compute the partial CI several times, each time by randomly drawing a set of 100 trials among the available 400, and we average the resulting CIs. Then we compute the absolute distance from this partial CI to the global CI. Values are normalized so that the farthest distance is 1 and the closest is 0 (corresponding to the global CI). The stability is finally derived by subtracting the distance from 1. Fig. 2 shows that, as stability comes close to 1, CI estimation is increasingly fair, at least according to the arbitrary reference of 400-trial CI. We also observe that stability is almost the same along the three parameters. If the value of 0.95 is defined as the accuracy threshold, it is reached around 260 stimuli, or 65% of the maximum value 400. Despite the fact that 260 is quite lower than 400, we may notice that the shape of the curve doesn't meet the usual shape of an hyperbola, suggesting that more stimuli could be necessary. We further examine this issue in the second experiment (see Section 4).

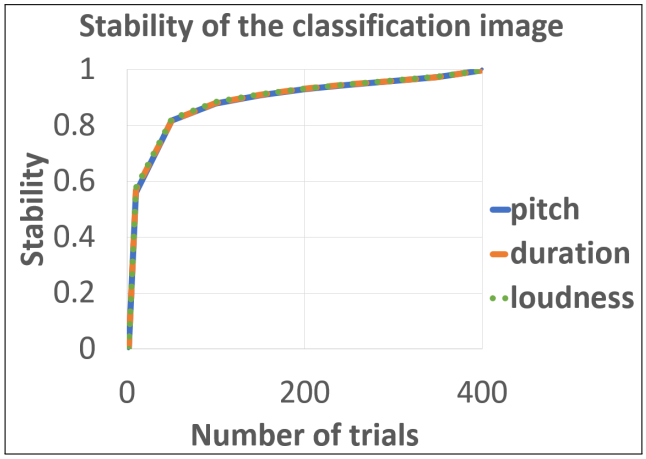

Figure 2. Stability of the CI of joy in speech in experiment 1, along its three parameters : pitch, duration and loudness.

Lastly, a kind of pairwise classification is performed, in order to compare the parameter values (i.e. the prosodic profile in terms of pitch, duration and loudness) between the stimuli selected as joyful and the discarded stimuli. The goal is to measure the relative weight of each feature in the observed judgments. The analysis is inspired from the hybrid filter-wrapper approach to feature selection described in [15] and is performed by the algorithm KHIOPS [16]. The performance of pairs of features is evaluated by a machine learning algorithm. Then the features are ranked on the basis of the matrix of pairwise performances numbers. In this way, the discriminatory ability of pairs of features is evaluated and compared. In our algorithm, the discriminatory ability is measured by using MODL criterion [17]. This analysis shows that pitch parameters are the most discriminant features and that the contribution of other features may be neglected. Particularly speaker gender and the pseudo-word have no impact.

4. EXPERIMENT 2

4.1 Presentation

Based on previous findings (see Section 3), we set up a new experiment, with the goal of, on one side, narrowing the scope of the experiment and improving the search when it is possible, and, on the other side, adding more variations, when it is necessary. We propose the following changes:

- we focus on pitch variations only;
- all the pitch variations are generated with a uniform distribution;
- variation range is larger, from -600 to 600 cents;
- no flattening is performed, variations are directly applied on the reference stimulus;
- only one speaker is considered (the male speaker);
- only one pseudo-word is considered (*bazin*);
- number of trials is raised to 1000 per subject, plus 100 repeated stimuli (divided into 22 blocks of 50 stimuli). Eleven blocks are presented in a first session, the remaining 11 are presented to the same subject in another session, a few days later.

For this exploratory experiment, we recruit 5 native French listeners (mean age = 39.8 +/-9,3 females). For each session, apparatus and procedure are identical to the previous experiment.

4.2 Results

The CI is computed in the same way as in the previous experiment. In Fig. 3, we observe that, similarly to experiment 1, the global shape of the curve starts from 0, increases in the middle, reaching a maximum peak at segment 3 (leading to a maximum pitch increase around the end of the first syllable), and goes back down until the end. For the repetition blocks, with the same process as in experiment 1, we get a mean self-consistency value of 68.2% +/-5.8, i.e., the exact same value with a smaller standard deviation.

4.3 Discussion

Despite a similar global 'increase then decrease' behavior between the pitch curves of the two experiments, we notice a few changes. First, the whole curve values are far greater in experiment 2 than in experiment 1. For instance, the peak value at segment 3 reaches 200.7 cents (i.e., 1 tone), instead of 28.8 cents in the previous experiment. Second, the peak value at segment 5 has now disappeared, even if the value is still bigger than in experiment 1 (70.9 cents here vs. 28.2 cents previously), it is lower than the other points of the curve. These differences between the two experiments reveal the influence of the modified method variables. However, although this pitch profile curve leads now to audible changes when reapplied to a reference stimulus, a proper validation phase is still

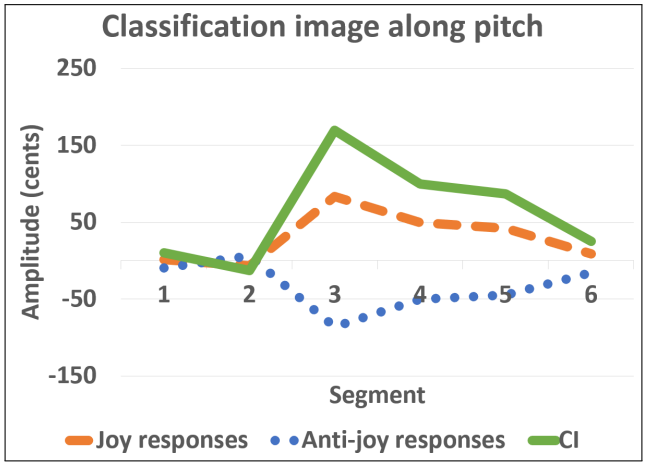

Figure 3. Average CI of joy in speech, in experiment 2, across all subjects, with pitch variations only. The medium-dashed line is the mean pitch of the stimuli classified as joyful, the small-dashed line is the mean pitch of the discarded stimuli (so-called "anti-joy"), and the solid line is the resulting CI.

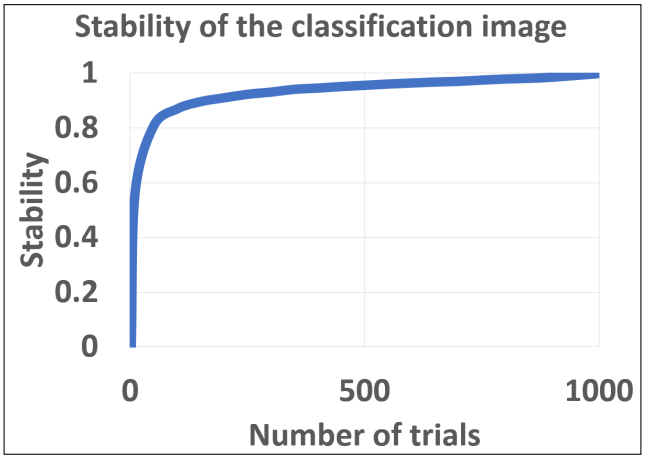

Figure 4. Stability of the CI of joy in speech in experiment 2, considering pitch variations only.

mandatory to check the validity of the CI.

Looking at CI stability (see Fig. 4), we observe that the threshold of 0.95 is now reached around 439 trials, or 43.9% of the maximum value 1000. Considering that the two experiments are subtly different, but close enough to be comparable, this suggests that 400 trials was indeed insufficient. Further experiments with different number of stimuli seems necessary to look for the threshold from which stability stops changing. Nevertheless, study of CI stability would benefit from additional metrics. For instance, an hyperbolic regression could be performed to predict the stability tendency for an increasing number of trials.

In this exploratory study, we pushed parameters in drastic directions, shrinking the stimulus space on one side (1 pseudo-word only, 1 speaker, considering pitch variations only) and expanding it on the other side (6 tones of amplitude for pitch variations, uniform distribution). Having obtained a quite different CI from experiment 1, it would be now interesting to methodically reintroduce complexity and observe the effects. In particular, one could question if duration and loudness profile curve values would benefit from the same increase with similarly expanded bounds.

5. GLOBAL DISCUSSION

Some recent efforts have been made to improve the reverse correlation method [18, 19]. This paper is in the wake of this movement. With our exploratory study, we raise questions in terms of how to conduct the search within the stimulus space (here the space of joy in speech): the properties of the base stimulus, with or without flattening, determine the anchor point in the space from where the exploration starts; the choice of parameters (pitch, duration, loudness) are the axes along which to explore; the choice of the random distribution relates to the heterogeneity of the search and where it stops; the number of stimuli defines the sampling frequency.

According to this point of view, the reverse correlation paradigm is analogous to an optimization problem: not only we are investigating the mapping between acoustic parameters and the mental representation of joy in speech, but we also want to identify where, in this mapping, the perception of joy is maximum, in order to derive a kind of "joy filter" which could transform a joyless speech into a joyful one in real time. From this perspective, we hope that our study can lead to further investigations on reverse correlation. For instance, to consider not only a global maximum, but also possible local maximums, and the possibility for a CI to get stuck in one of them. One could also study temporal dependencies of a parameter, in addition to dependencies across parameters.

6. CONCLUSION

We proposed two experiments implementing the reverse correlation method, to capture mental representation of joy in speech regarding three acoustic parameters: pitch, duration and loudness. These experiments aimed to compare, in an exploratory way, the effect of several protocol variables on the resulting CIs. We observed noticeable differences, leading us to outline the remaining work needed before being able to merge the two notions of CI and mental representation. Further studies should carry on this prospect, to gradually establish a more consistent link between reverse correlation variables and resulting CIs.

7. REFERENCES

[1] J. Eggermont, P. Johannesma, and A. Aertsen, "Reverse-correlation methods in auditory research," *Quarterly reviews of biophysics*, vol. 16, no. 3, pp. 341–414, 1983.

[2] M. Kim, J. M. Gold, and R. F. Murray, "What image features guide lightness perception?," *Journal of vision*, vol. 18, no. 13, pp. 1–1, 2018.

[3] M. C. Mangini and I. Biederman, "Making the ineffable explicit: Estimating the information employed for face classifications," *Cognitive Science*, vol. 28, no. 2, pp. 209–226, 2004.

[4] R. E. Jack, R. Caldara, and P. G. Schyns, "Internal representations reveal cultural diversity in expectations of facial expressions of emotion.," *Journal of Experimental Psychology: General*, vol. 141, no. 1, p. 19, 2012.

[5] L. Wang, J. H. Ong, E. Ponsot, Q. Hou, C. Jiang, and F. Liu, "Mental representations of speech and musical pitch contours reveal a diversity of profiles in autism spectrum disorder," *Autism*, p. 13623613221111207, 2022.

[6] L. Goupil, E. Ponsot, D. Richardson, G. Reyes, and J.-J. Aucouturier, "Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature," *Nature communications*, vol. 12, no. 1, p. 861, 2021.

[7] E. Ponsot, J. J. Burred, P. Belin, and J.-J. Aucouturier, "Cracking the social code of speech prosody using reverse correlation," *Proceedings of the National Academy of Sciences*, vol. 115, no. 15, pp. 3972–3977, 2018.

[8] E. Ponsot, P. Arias, and J.-J. Aucouturier, "Uncovering mental representations of smiled speech using reverse correlation," *The Journal of the Acoustical Society of America*, vol. 143, no. 1, pp. EL19–EL24, 2018.

[9] L. Brinkman, A. Todorov, and R. Dotsch, "Visualising mental representations: A primer on noise-based reverse correlation in social psychology," *European Review of Social Psychology*, vol. 28, no. 1, pp. 333–361, 2017.

[10] F. Gosselin and P. G. Schyns, "Superstitious perceptions reveal properties of internal representations," *Psychological science*, vol. 14, no. 5, pp. 505–509, 2003.

[11] L. Brinkman, S. Goffin, R. van de Schoot, N. E. van Haren, R. Dotsch, and H. Aarts, "Quantifying the informational value of classification images," *Behavior Research Methods*, vol. 51, pp. 2059–2073, 2019.

[12] C. Sobin and M. Alpert, "Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy," *Journal of psycholinguistic research*, vol. 28, pp. 347–365, 1999.

[13] M. Schröder, "Emotional speech synthesis: A review," in *Seventh European Conference on Speech Communication and Technology*, 2001.

[14] J. J. Burred, E. Ponsot, L. Goupil, M. Liuni, and J.-J. Aucouturier, "Cleese: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition," *PloS one*, vol. 14, no. 4, p. e0205943, 2019.

[15] S. Dreiseitl and M. . Osl, "Feature selection based on pairwise classification performance," in *Computer Aided Systems Theory - EUROCAST 2009*, 2009.

[16] `www.khiops.com`. Accessed: 2023-04-21.

[17] M. Boullé, "Modl: a bayes optimal discretization method for continuous attributes," *Machine learning*, vol. 65, pp. 131–165, 2006.

[18] M. Kevane and B. Koopmann-Holm, "Improving reverse correlation analysis of faces: Diagnostics of order effects, runs, rater agreement, and image pairs," *Behavior Research Methods*, pp. 1–39, 2021.

[19] A. Compton, B. W. Roop, B. Parrell, and A. C. Lammert, "Stimulus whitening improves the efficiency of reverse correlation," *Behavior research methods*, pp. 1–9, 2022.