DOI: 10.61782/fa.2023.0268

AUDITORY REVERSE CORRELATION APPLIED TO THE STUDY OF PLACE AND VOICING: FOUR NEW PHONEME-DISCRIMINATION TASKS

Géraldine Carranante[1*] **Maria Giavazzi**[2] **Léo Varnet**[1]
[1] LSP, ENS, PSL University, CNRS, 75005 Paris, France
[2] ENS, PSL University, CNRS, 75005 Paris, France

ABSTRACT

Auditory reverse correlation (revcorr) is an experimental paradigm that reveals the acoustic cues used by listeners in any auditory task. It has been previously used to explore the categorisation of /aba/ and /ada/ sounds in noise. Here, we extend the paradigm to new phonemic contrasts. In a typical revcorr experiment, one introduces random fluctuations in stimuli in order to measure how they affect the behavioural responses of the participant on a trial-by-trial basis. The outcome is called auditory classification images (ACI), i.e. time-frequency maps of the acoustic cues used by participants, revealing their individual listening strategies in a given task. Here, we use the "fastACI toolbox" [Osses & Varnet, 2021] to apply the paradigm to new phonemic contrasts : /aba/-/apa/ ; /ada/-/aga/ ; /ada/-/ata/ ; /apa/-/ata/. It allows us to study the perception of two phonetic traits: place of articulation and voicing. We present the results of 2 participants for each contrast. The results are consistent with the main auditory cues already identified in the psycholinguistic literature but they also reveal unexpected secondary cues.

Keywords: *phoneme categorisation, reverse correlation, auditory classification image, phonetic traits*

*Corresponding author: geraldine.carranante@ens.fr.

1. INTRODUCTION

Understanding speech is a complex process that requires, as a first step, the decoding of phonemes or syllables from the acoustic input. To achieve this goal, the human brain must rely on the detection of specific acoustic features, known as "acoustic cues", that are characteristic of each phoneme. These cues help to distinguish one phoneme from another and allow the accurate decoding of speech sounds. In this project, we aim to identify the acoustic cues that native French speakers rely on to recognize French plosive consonants (/p/, /t/, /k/, /b/, /d/, /g/), and to improve our understanding of the low-level auditory mechanisms underlying speech perception.

The identification of acoustic cues that are critical for speech perception has been a long-standing problem in psycholinguistics, mainly because the spectro-temporal complexity of natural speech makes it difficult to isolate specific cues and to manipulate them independently. Additionally, speech signals are redundant, providing a large number of possible, co-varying cues to the listener. Researchers have used various psycholinguistic methods to identify the acoustic cues used in speech perception. These approaches make use of "reduced" speech stimuli to isolate the perceptual effects of individual acoustic cues. In particular, previous studies have relied on high-pass and low-pass filtered speech [1, 2], time-truncated speech [2], or speech synthesized with only a limited number of cues [3–5]. Applied to the study of plosive consonant perception, these approaches have demonstrated that place-of-articulation contrasts (i.e. /b/-/d/-/g/, /p/-/t/-/k/) primarily involve the detection of release bursts and F2 onset cues. On the other hand, voicing contrasts cues (i.e. /b/-/p/, /d/-/t/, /g/-/k/) were found to include the Voice Onset Time (VOT), f_0 onset frequency, release burst intensity, and the relative duration of the vocalic segments.

These results were obtained using stimuli artificially manipulated to isolate the perceptual effects of individual acoustic cues. As a result, they may not fully capture the complexity of natural speech. Indeed, the speech comprehension system can adapt its strategy in response to signal reductions, meaning that the cues identified with these approaches may not necessarily reflect the cues used during natural speech perception.

To address this limitation, the present study aims to investigate the perception of plosive consonants based on natural speech utterances. For this purpose, we use an auditory reverse correlation (auditory revcorr) approach to identify the specific acoustic cues used by listeners in a phoneme-discrimination task. Originally proposed by Ahumada and Lovell [6], the auditory revcorr method has become very popular in the psychoacoustic community during the last decade. In the present study, we focus on a specific version of the auditory revcorr paradigm particularly suited for exploring phoneme-discrimination tasks and called auditory classification image (ACI).

In a typical ACI experiment, participants are asked to discriminate two sounds that are repetitively presented in a background noise. This paradigm, which corresponds to a very simple form of sound categorisation, is illustrated in Fig. 1 (grey boxes). In each trial, one speech target (here, /ada/ or /aga/) is chosen at random and embedded in additive noise. After each stimulus presentation, the participant is asked to indicate which target they heard. In some trials, the random distribution of noise will mask crucial acoustic characteristics of the target, resulting in an incorrect response of the participant. The rationale behind the ACI approach is to find the systematic relationship between the exact time-frequency (T-F) configuration of the noise in each trial and the corresponding response of the listener (blue box in Fig. 1). This statistical relationship between the masker and the phonetic response is summarized as a matrix of perceptual weights, the ACI, which reveals the T-F regions where the presence of noise misled the participant in a systematic way. These regions correspond to the location of the acoustic cues the participant relied on to discriminate the two targets. The ACI method can therefore be described as a perceptual imaging technique as it provides a way to visualise the listening strategy used by a participant during the task. This novel methodology is based on purely behavioural data (no neuroimaging data) and provides an unprecedented insight into the mechanisms at play at the acoustic-phonetic

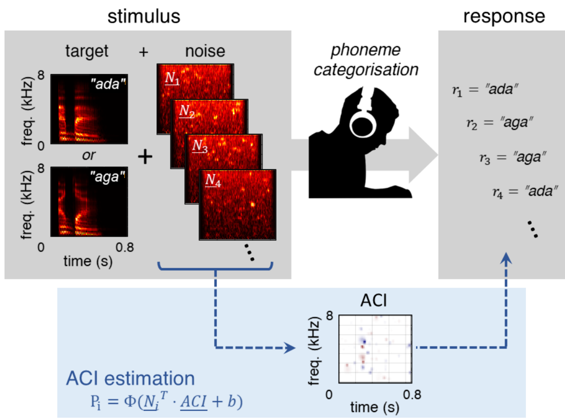

Figure 1. Schematic diagram of a typical ACI experiment (grey boxes) and ACI estimation (blue boxes). See text for details.

interface. The fastACI toolbox developed in our group and freely available online, offers a ready-to-use solution for setting up, running, and analysing an ACI experiment in a MATLAB environment [7].

The ACI approach has been previously used to explore place-of-articulation perception using different pairs of “aba” and “ada” utterances [8–10]. In this short communication, we outline an extension of this work to four new phonemic contrasts: /b/-/p/, /d/-/g/, /d/-/t/ and /p/-/t/.

2. METHODS

This study presents data from 2 participants (S1 and S13) on 5 experimental conditions. The participants are two of the authors of this study, aged 35 and 33 years old, 1 male and 1 female, both native French speakers, with at least one normal hearing ear (validated by a standard audiometric procedure, with a threshold equal or lower than 20dB HL). Each condition corresponds to a pair of target vowel-consonant-vowel sounds: (1) **ABDA22** condition contains the pair /aba/-/ada/, (2) **ADGA23** condition contains the pair /ada/-/aga/, (3) **ABPA23** condition contains the pair /aba/-/apa/, (4) **ADTA23** condition contains the pair /ada/-/ata/ and (5) **APTA23** condition contains the pair /apa/-/ata/. Each target sound is presented to the participant in noise on single-interval trials. Data on ABDA22 condition for S1 were collected as part of a previous study [9]. Data for S13 on ABDA22 condition as well as all the other conditions for both subjects are original data. The experiment was conducted in the two double-walled soundproof booths at ENS Paris and can be reproduced within the fastACI toolbox (see Listing 1).

2.1 Experimental protocol

The experiment is based on a discrimination task in noise with a pair of equally-likely targets presented diotically via headphones (Sennheiser HD 650 circumaural headphones). Each condition consisted in 4000 trials, with each trial having one target-in-noise interval to which the participant had to indicate one of the two possible answers on a keyboard. The two target sounds were presented 2000 times in each condition, in a randomized order. The structure of a single trial is shown in Fig. 1 (for the ADGA23 condition). The same protocol is used for all conditions, with different target sounds. For each condition, the 4000 trials were divided in 10 blocks of 400 trials, during which the signal-to-noise ratio was adapted following a one-up one-down weighted adaptation rule, with a ration of 2.41 to 1, to ensure a mean performance of 70.7%. The change of signal-to-noise (SNR) is implemented in adjusting the level of the target sound. After the first block of 400 trials for each condition, the SNR threshold is calculated. A SNR threshold over -11.5 dB leads to the exclusion of the participant from this condition.

Listing 1. MATLAB commands required to reproduce the experiment described in this paper.

```
cond = 'abda'; % condition name ('aCCa')
subj_ID = 'S01'; % subject ID
exp_type = 'speechACI_Logatome';
noise_type = 'bumpv1p2_10dB';
exp_name = [exp_type '-' cond '-S43M'];
fastACI_experiment(exp_name,subj_ID,noise_type);
```

2.2 Target stimuli and noise

Target sounds are 5 logatomes uttered by male speaker S43M taken from the OLLO speech corpus [11], listed in Table 1. The speech samples were preprocessed to align the time position of the vowel-consonant transitions, and to equalize the duration and energy of each syllable. The equalized sounds have a sampling frequency of 16000Hz, an overall duration of 0.86s and a level of 65 dB SPL. The time-frequency representation of the preprocessed sounds is presented in Fig. 4, together with their fundamental-frequency (f_0) and formant $(F_1 - F_4)$ trajectories. The darker and lighter regions represent T-F points with respectively lower and higher amplitudes.

Table 1. Original target sounds from the OLLO speech corpus

Logatome	Filename
/aba/	S43M_L007_V6_M1_N2_CS0.wav
/ada/	S43M_L001_V6_M1_N1_CS0.wav
/aga/	S43M_L003_V6_M1_N2_ CS0.wav
/apa/	S43M_L008_V6_M1_N1_CS0.wav
/ata/	S43M_L002_V1_M1_N1_CS0.wav

The background noise, called ”bump noise”, has a spectral content similar to a white noise, but with stronger envelope fluctuations below 35 Hz [9]. These low-frequency fluctuations were generated by superimposing bumps to a white noise. These bumps are highly-energetic Gaussian-shaped regions, with a temporal width of $\sigma_t = 0.02s$, a spectral width of $\sigma_f = 0.5$ ERB and an amplitude emphasized to max 10 dB. For each participant and condition, a new set of 4000 noises with the same sampling frequency and duration as the target sounds are generated.

In each trial, the level-adjusted target sounds are arithmetically added to the noises to produce noisy speech sounds. The presentation level is additionnaly slightly modulated between -2.5 to +2.5 dB (continuous range, uniform distribution).

2.3 Analysis

2.3.1 Time-frequency noise representations

Each noise waveform is converted into a time-frequency (T-F) representation, using a Gammatone-based representation as in [9, 10, 12]. The noises are thus decomposed into 64 bands equally spaced in the ERB scale between 45.8 Hz and 8000 Hz. Then, the 64 band-passed signal are low-pass filtered to simulate approximately an inner-hair-cell envelope processing. Finally, the T-F representation was reduced to a 86-by-64 matrix, by calculating one estimate (amplitude mean) every 0.01s for each frequency band. This matrix is reshaped into a 5504-by-1-elements vector $\underline{N_i}$.

2.3.2 Statistical estimation

The revcorr approach consists in measuring the impact of random fluctuations within stimuli on the behavioral responses of the participants on a trial-by-trial basis [13].

In order to reveal the statistical relationship between the stimulus presented in trial i and the participant's response r_i, the revcorr approach relies on a stimulus-response transformation based on a generalised linear model (GLM, blue box in Fig. 1). The model output is a predicted probability P_i that the participant gave the response 1. For example, in condition ABDA22, $P_i = P(r_i = \text{"aba"}) = 1 - P(r_i = \text{"ada"})$. The T-F information from the noise is combined linearly using a matrix of decision weights, called an auditory classification image (ACI), such that:

$$P_i = \Phi(\underline{N_i}^T \cdot \underline{\text{ACI}} + b) \quad (1)$$

The GLM parameters (i.e., the vectorised matrix $\underline{\text{ACI}}$ and the scalar intercept b) are fitted individually based on each participant's data, such that the logistic function Φ returns a P_i value closer to 1 or 0 when the participant selected response 1 or 2, respectively.

Each $\underline{\text{ACI}}$ element can be interpreted as a perceptual weight in favour of one of the two answers, associated to one particular T-F point. The matrix form $\underline{\underline{\text{ACI}}}$ has therefore the same format as the T-F representations of the target sounds, with coloured regions representing these perceptual weights towards one of the responses (see Fig. 4). It provides a direct way of assessing the listening strategy of the participant in the task, i.e., the T-F location of the cues they rely on.

The GLM is fitted with a L1 (lasso) penalty on a Gaussian Pyramid basis. This type of penality relies on the assumption that the cues are sparse in the T-F space, and can be decomposed into a sum of Gaussian elements [9, 14]. It therefore ensures that weights in the ACI which do not strongly improve the GLM predictions are suppressed, while critical weights, useful to explain the data, are emphasized.

3. RESULTS AND DISCUSSION

The overall behavioral performances for each participant and each condition are provided by the measure of the SNR threshold yielding 70.7% of correct responses in the task. The SNR thresholds per condition for each participant are shown in Fig. 2 and are very similar, ranging between -15.7 and -12.3 dB (the lower the better). They show that both participants are able to perform the task. It is however important to note that these levels of SNR are very low (the noisy sounds are very noisy), which limits the naturalness of the task.

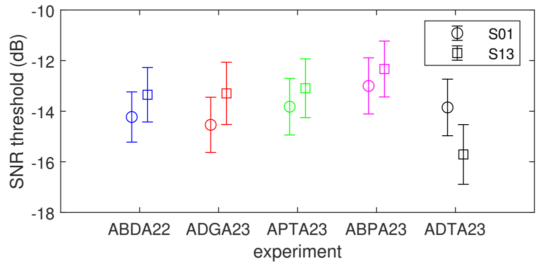

Figure 2. SNR threshold for the two participants in each of the five conditions. Error bars indicate ± 0.5 SD

The left column of Fig. 3 shows the sensitivity for each condition and each participant as a function of SNR. In Signal Detection Theory, the sensitivity (d') is a measure of the perceptual distance between two signals [15]. Compared to percentage of correctness, it allows to disentangle the ability to detect a signal from biases of participants toward an answer. The latter is expressed by a criterion value (c), shown in the right column of Fig. 3. As expected here, we observe a lower sensitivity when the SNR decreases, i.e. when the task becomes more difficult. The psychometric functions (sensitivity as a function of SNR) are very similar across all conditions and both participants.

The biases towards one of the answers in all conditions, shown in the right column of Fig. 3, are quite low, especially for subject S13. It has however been shown in another study [9] that there is a systematic dependance of bias to SNR. Further investigations are needed on more subjects to determine the effect of bias in this type of task.

The ACIs of participants S01 and S13 for each condition are shown in the third and fourth columns of Fig. 4. The ACIs show distinct weighted regions, corresponding to T-F configurations of noise energy that result in a higher probability that the participant gives one of the two possible answers. Note that the association of positive and negative weights (red and blue weights, respectively) to a specific target in each contrast is arbitrary. For the 3 conditions of place-of-articulation perception (ABDA22, ADGA23 and APTA23), positive values are associated with the more front consonant; for the 2 conditions of voicing perception (ABPA23 and ADTA23), positive values are associated with the voiced consonant. To interpret the results of Fig. 4, the ACIs need to be mentally "superimposed" on the spectrograms of the target sounds pre-

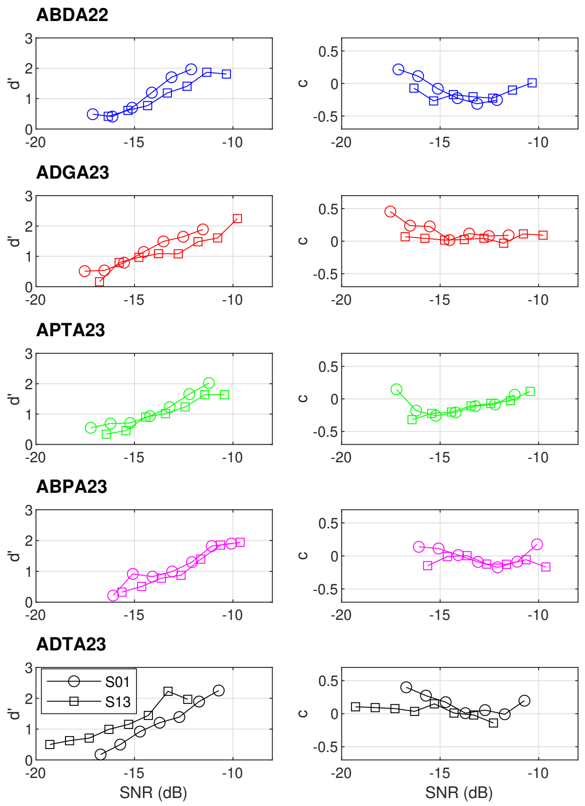

Figure 3. Performance and bias as a function of SNR for the two participants in each of the five conditions. Performance is measured as a sensitivity (d') value while bias is reported as a criterion (c) value.

sented on the two leftmost columns of the figure, in order to find the specific acoustic cues these weights correspond to. The two groups of conditions (place-of-articulation conditions and voicing conditions) show two specific patterns of weights, that we describe in the next two paragraphs.

3.1 Place-of-articulation perception

The ACIs corresponding to place-of-articulation contrasts (ABDA22, ADGA23, APTA23) reveal strong weights located at the onset of the second syllable (t=0.3s). These weights show a similar pattern centered around the F2 onset frequency ($\sim$ 1600 Hz), with a positive (red) weight below a negative (blue) weight. The position of these two weights approximately matches the position of the F2 onset for the corresponding targets (1298 Hz for /aba/, 1722 Hz for /ada/, 2080 Hz for /aga/, 1300 Hz for /apa/, 1616 Hz for /ata/). Therefore, the presence of noise energy at the T-F location of one of the two targets' F2 onset frequency biases the perception towards this target. This finding confirms results from previous studies which used synthetic speech [3, 4], showing that the F2 onset is a predominant cue in natural speech perception in noise.

Similarly to the F2 onset, the first three experimental conditions also yield weak but consistent weights in the region of the F1 onset, indicating that the exact frequency position of this formant is also a cue for place-of-articulation perception in noise. References to a possible role of F1 transitions in place perception are more seldom [5, 16]. However, this cue was already found in previous ACI studies on /aba/-/ada/ [9] and /alda/-/alga/-/arda/-/arga/ [17] categorization tasks.

Finally, all ACIs measured for participant S13 for place contrasts (conditions ABDA22, ADGA23, APTA23) show strong high-frequency weights at syllable onset. This reflects the use of a consonant burst release cue for the detection of dental plosives /d/ and /t/ [2]. The importance and relative weights of F2-transition cues and burst cues in the perception of plosive consonants has already been discussed in length [2, 5, 16]. The present results suggest that interindividual variability may play an important role in the balance between these two cues, with participant S01 not relying on burst cues and participant S13 consistently using them across 3 conditions.

3.2 Voicing perception

The primary cue for the two voicing contrasts (ABPA23 and ADTA23) is located between $\sim$ 0.2s-0.3s in the very low frequencies (around $\sim$ 200 Hz), which coincides with the location of the so-called "voicing bar" (corresponding roughly to the energy in the first and second harmonics) during the intervocalic interval. The horizontal configuration of the weights, with negative weights preceding positive weights, confirms that the critical characteristic of this cue is its temporal position. Thus, these weights reflect the use of a VOT cue, with more low-frequency noise energy during the intervocalic interval being interpreted as a more voiced phoneme (positive weights). By contrast, a more energetic syllable before the intervocalic interval reinforces the opposite percept, yielding more unvoiced responses (negative weights).

The ACIs measured for the two voicing contrast conditions also exhibit weak positive and negative weights at the F1 onset and offset location, consistent with the use of a F1 cue for voicing perception. Although it is well-known that voiced plosive consonants have lower F1 onsets [18], there are limited experimental investigation of the perceptual role of this cue [19]. The current findings suggest that listeners are able to extract and use this information when engaged in a voicing discrimination task.

Other cues for voicing perception have been reported in previous psycholinguistic works, including the relative duration of the vocalic segments, the f_0 onset frequency and the release burst intensity [18, 20]. However, these cues were not found in the measured ACIs. One possible explanation for this discrepancy is the very low SNR under which the task was performed. In such challenging listening conditions, the participants might have adapted their perceptual strategies to focus on the most robust cues.

As the two participants S01 and S13 are also authors of this study, it is important to consider to what extend their familiarity with the experiment may have influenced the results. This familiarity has two aspects: first, both participants have a lot of practice with the task, i.e., an acquaintance with the stimuli and a procedural learning of the experimental set up. Second, they have expert knowledge about phonetics. With respect to the first type of familiarity, previous experiments on 12000 trials [9] indicated no effect of learning throughout the course of the experiment. Concerning an alleged expertise effect, the same study mixed expert participants with naïve ones with regards to the task, and no performance benefit was shown for the former group. Therefore, the fact that the two participants are the authors should not have major influence on the outcome of this study. This will be confirmed in the follow-up of the present study, which will include a larger number of participants in each condition.

4. CONCLUSION

Psycholinguists have employed a variety of experimental methods to investigate the role of different acoustic cues in phoneme categorization, including filtering, time-truncation, and speech synthesis. An argument in favor of these manipulation-based experiments is that the speech features can be carefully controlled. However, the major disadvantage is that they require prior knowledge of the specific cues that can play a role in the decoding process and, subsequently, they cannot probe the large space of all possible cues available to the listeners. In contrast, the ACI method provides a comprehensive approach to explore speech cues, allowing the investigation of many acoustic dimensions simultaneously with only minimal assumptions.

The results presented in this paper suggest that the ACI method may be effective in identifying cues for place of articulation and voicing in stop consonant recognition. The ACIs measured in two participants confirmed the role of F2 transitions, VOT and release burst cues, already documented in previous studies. They also indicate a potential role of F1 transitions — a result which has yet to be confirmed on a larger group of listeners. Finally, the two participants show consistent differences in their listening strategies, in particular in the weight attributed to the high-frequency burst cue. This heterogeneity can be quantified and assessed within the ACI framework. This will be the object of a future article.

Understanding interindividual differences in speech listening strategies is instrumental for the development of more effective hearing aid systems, as well as more refined diagnostic tools. Hearing aid professionals often face significant individual variation when it comes to the effectiveness of fitted hearing aids in their patients. This variability is related to the "hidden" components of hearing loss and possibly to differences in individual listening strategies. As a consequence, it is crucial to customize and fine-tune hearing aid systems for each individual listener. In this regard, the ACI framework offers an objective means of assessing individual listening strategies, enabling a truly personalized adjustment of corrective systems provided to individuals with hearing impairment. This will be the focus of future research from our group.

5. DATA AVAILABILITY

The fastACI toolbox for MATLAB, containing all the required scripts to replicate the results from this study can be downloaded from `https://github.com/aosses-tue/fastACI`. All figures can be replicated using the fastACI script `publ_carranante2023_FA_figs.m`.

6. ACKNOWLEDGMENTS

This study was funded by the ANR grants "fastACI" (Grant No. ANR-20-CE28-0004), "DIA-SYN-PHON" (Grant No. ANR-21-CE28-0008) and "FrontCog" (Grant No. ANR-17-EURE-0017).

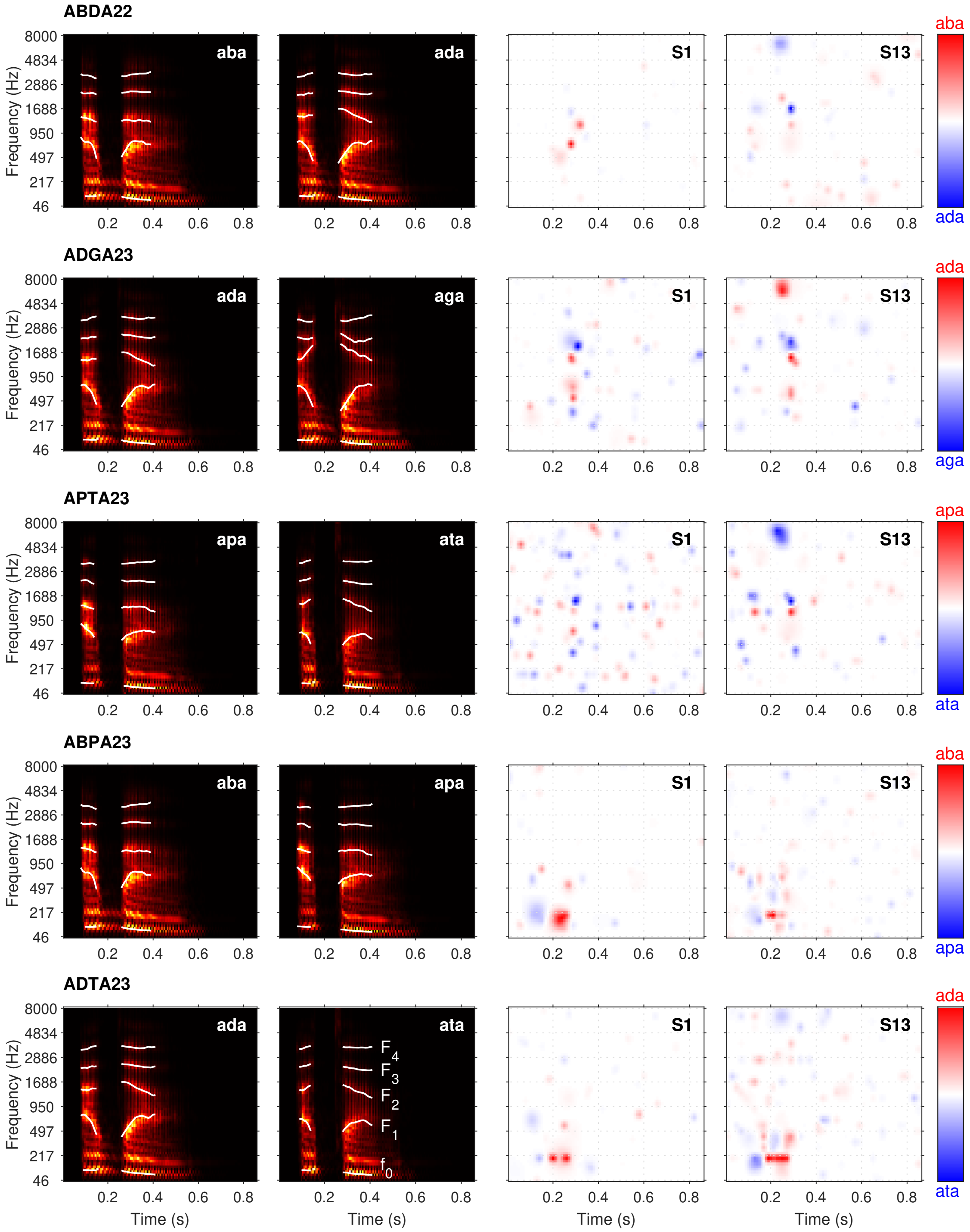

Figure 4. Target sounds and ACIs. The two leftmost columns display the target sounds used in the five conditions, with the formants and fundamental frequency indicated with white lines. The corresponding ACIs obtained by the two participants are shown in the two rightmost columns.

7. REFERENCES

[1] H. Fletcher, "The nature of speech and its interpretation," *Journal of the Franklin Institute*, vol. 193, pp. 729–747, June 1922.

[2] F. Li, A. Menon, and J. B. Allen, "A psychoacoustic method to find the perceptual cues of stop consonants in natural speech," *The Journal of the Acoustical Society of America*, vol. 127, pp. 2599–2610, Apr. 2010.

[3] A. M. Liberman, P. C. Delattre, and F. S. Cooper, "The role of selected stimulus-variables in the perception of the unvoiced stop consonants," *The American journal of psychology*, vol. 65, pp. 497–516, Oct. 1952.

[4] A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman, "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," *Psychological Monographs: General and Applied*, vol. 68, no. 8, pp. 1–13, 1954.

[5] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," *The Journal of the Acoustical Society of America*, vol. 27, pp. 769–773, June 1955.

[6] A. J. Ahumada, Jr and J. Lovell, "Stimulus Features in Signal Detection," *The Journal of the Acoustical Society of America*, vol. 49, no. 6B, pp. 1751–1756, 1971.

[7] A. Osses and L. Varnet, "fastACI toolbox: the MATLAB toolbox for investigating auditory perception using reverse correlation (v1.0)," Oct. 2021.

[8] L. Varnet, K. Knoblauch, F. Meunier, and M. Hoen, "Using auditory classification images for the identification of fine acoustic cues used in speech perception.," *Frontiers in Human Neuroscience*, vol. 7:, p. 865, 2013.

[9] A. Osses and L. Varnet, "A microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception," Dec. 2022. Pages: 2022.12.27.522040 Section: New Results.

[10] A. Osses and L. Varnet, "Consonant-in-noise discrimination using an auditory model with different speech-based decision devices," in *DAGA*, 47th Annual Conference on Acoustics DAGA 2021 Wien, (Vienne, Austria), pp. 298–301, Aug. 2021.

[11] T. Wesker, B. Meyer, K. Wagener, J. Anemüller, A. Mertins, and B. Kollmeier, "Oldenburg logatome speech corpus (OLLO) for speech recognition experiments with humans and machines," in *in Proceedings of Interspeech*, pp. 1273–1276, 2005.

[12] A. Osses and L. Varnet, "Auditory Reverse Correlation on a Phoneme-Discrimination Task: Assessing the Effect of Different Types of Background Noise," Feb. 2022. Published: ARO 2022.

[13] R. F. Murray, "Classification images: A review," *Journal of Vision*, vol. 11, Jan. 2011.

[14] P. J. Mineault, S. Barthelmé, and C. C. Pack, "Improved classification images with sparse priors in a smooth basis," *Journal of vision*, vol. 9, no. 10, pp. 17.1–24, 2009.

[15] L. O. Harvey, "Detection Theory: Sensitivity and Response Bias," 2004.

[16] A. Alwan, J. Jiang, and W. Chen, "Perception of Place of Articulation for Plosives and Fricatives in Noise," *Speech communication*, vol. 53, pp. 195–209, Feb. 2011.

[17] L. Varnet, K. Knoblauch, W. Serniclaes, F. Meunier, and M. Hoen, "A Psychophysical Imaging Method Evidencing Auditory Cue Extraction during Speech Perception: A Group Analysis of Auditory Classification Images," *PLoS ONE*, vol. 10, p. e0118009, Mar. 2015.

[18] L. Lisker, ""Voicing" in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees," *Language and Speech*, vol. 29, pp. 3–11, Jan. 1986. Publisher: SAGE Publications Ltd.

[19] J. Benki, "Place of articulation and first formant transition pattern both affect perception of voicing in English," *Journal of Phonetics*, vol. 29, pp. 1–22, Jan. 2001.

[20] J. Schertz and E. J. Clare, "Phonetic cue weighting in perception and production," *WIREs Cognitive Science*, vol. 11, no. 2, p. e1521, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1521.