INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

No distributional learning in adults from attended listening to non-speech

Ellen Marklund, Johan Sjons, Lisa Gustavsson, Elísabet Eir Cortes

Stockholm Babylab, Phonetics Laboratory, Department of Linguistics, Stockholm University
`ellen.marklund|johan.sjons|lisa.gustavsson|elisabet.cortes@ling.su.se`

Abstract

Distributional learning is a perceptual process hypothesized to underlie the phenomena of phonetic recalibration and selective adaptation, as well as infant speech sound category learning. However, in order to be conclusively tied to the earliest stages of speech sound category development, that is, the formation of novel perceptual categories, distributional learning must be shown to operate on stimuli for which there are no pre-existing categories. We investigated this in a previous study, finding no evidence of distributional learning in adults from unattended listening to non-speech. Since attention to stimuli impacts distributional learning, the present study focused on distributional learning from attended listening to non-speech. The same paradigm was used as in the previous study, except that participants' attention was directed towards stimuli by means of a cover task. Non-speech stimuli were spectrally rotated vowels and the mismatch negativity was used to measure perceptual categorization. No distributional learning was found, that is, no effect of attention on distributional learning was demonstrated. This could mean that the distributional learning process does not operate on stimuli where perceptual categories do not already exist, or that the mismatch negativity measure does not capture the earliest stages of perceptual category development.

Index Terms: distributional learning, speech sound category development, spectrally rotated speech, MMN

1. Introduction

Distributional learning (DL) is a perceptual process hypothesized to underlie infant speech sound category learning [1], as well as the phenomena of phonetic recalibration [2] and selective adaptation [3]. The DL process entails the distributional properties of recent input impacting perceptual categorization, and is typically assessed by discrimination before and after exposure to sounds varying along an acoustic continuum, presented according to either a unimodal (one-category) or a bimodal (two-category) frequency distribution (Figure 1). Shifting of speech sound category boundaries has been demonstrated experimentally through DL both in infants [4], [5], [6], [7], [8], [9], [10] and in adults [11], [12], [13], [14], [15].

However, in order to be conclusively tied to the earliest stages of speech sound category development, that is, the formation of novel perceptual categories, DL must be shown to operate on stimuli for which there are no pre-existing categories. This study therefore focuses on testing whether DL takes place for non-speech sounds that have the same complexity as speech sounds.

In a study investigating perceptual categorization of complex non-speech sounds, no evidence for DL was found [16]. However, the participants in that study were instructed not to pay attention to the stimuli. Typically, in DL studies with speech stimuli, participants are asked to listen carefully to the stimuli (*e.g.*, [12], [13]), and it has been shown that attention has an effect on DL. Ong and colleagues found DL of lexical tone in adults only when the stimuli were attended [15], and in 10-month-old infants DL was found only in the group of infants that was most attentive to the stimuli [17].

The present study thus investigates whether attention to the signal results in DL of complex non-speech in adults. Identical experimental setup as in [16] was used, except that participants' attention was directed towards the acoustic stimuli during exposure, through a cover task (*cf.* [15]). Perceptual categorization was assessed before and after exposure by means of mismatch negativity (MMN) amplitude, since MMN has been shown to be a probe of perceptual categorization [18]. During the MMN blocks, attention was still directed away from the stimuli in line with customary practice [19]. If DL occurs during exposure, the MMN amplitude is expected to increase in the bimodal group and/or decrease in the unimodal group.

2. Method

2.1. Participants

Participants were 16 adults between 20 and 55 years (mean age 30 years, SD = 11). Half of the participants were assigned to the unimodal group and the other half were assigned to the bimodal group. Eight additional subjects participated in the study but were excluded due to technical problems during data collection (n = 4) or experimenter error (n = 4). Participants' first language(s) varied. Since stimuli were non-speech, a specific native language or number of native languages were not a criterion for participation. Right-handedness was not a criterion for inclusion as the MMN is measured in frontocentral areas, and no laterality effects are expected, but the majority of participants were right-handed (88%). Participants gave informed consent before taking part in the experiment, and received two movie vouchers as thanks for their participation. The study has been approved by the regional Ethical Review Board (2015/63-31).

2.2. Stimuli

A vowel continuum was synthesized from recordings of the two Swedish vowels /e/ and /i/, and each token was then spectrally rotated. For details on stimuli creation, see [16]. Spectrally rotated speech is of comparable acoustic complexity as speech, but not identifiable as speech [20], [21]. Importantly, it can be assumed that no learned perceptual categories exist in adults for rotated speech, as they have most likely never been exposed to it previously. All rotated vowels, R1 to R8 (see Figure 1), were presented during the exposure phase. In the MMN blocks, R3 was used as standard stimulus and R6 was used as deviant stimulus. Stimuli token duration was 340 ms.

http://dx.doi.org/10.21437/Interspeech.2019-1674

2.3. Experiment design

There were three blocks in the experiment: one pre-exposure MMN block, one exposure block and one post-exposure MMN block. The MMN blocks were identical and were comprised of 1000 trials each, 80% of which were standards and 20% of which were deviants. In the exposure block, 320 stimuli from along the full continuum were presented, according to different frequency distributions for the two groups of participants (see Figure 1). Among those stimuli, 40 sine tones were interspersed at random intervals. Participants were given the task to mark which stimuli were tones (on a list with the numbers 1-360), in order to direct their attention to the acoustic signal during exposure. For further details on the experiment, see [16].

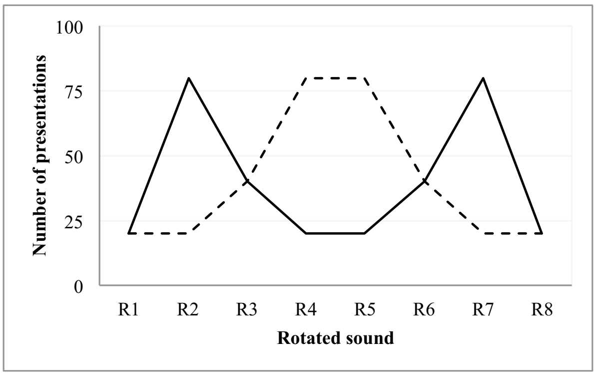

Figure 1: *The frequency distributions of the rotated vowels presented during exposure. The dashed line shows the distribution presented to participants in the unimodal group and the solid line shows the distribution presented to participants in the bimodal group.*

2.4. Procedure

Participants filled out a consent form and selected a movie to watch during the MMN blocks, each lasting approximately 12 minutes. They were instructed on their tasks (to be performed during the experiment) while the EEG head-cap was applied. In order to optimize data quality, participants were asked to remain as still as possible during stimuli presentation in the MMN blocks, and to move around, blink and stretch during brief pauses that occurred approximately every two minutes. During the MMN blocks, the participants watched their selected movie with Swedish subtitles but no sound. Before the start of the exposure block, both the movie and stimulus presentation paused. Participants were reminded that their task during the middle part of the experiment was to mark which stimuli were tones, among the other sounds, and given a pen and a sheet of paper with numbers on it, in order to be able to do so. Once the participants were ready, stimuli presentation continued throughout the exposure block, which lasted approximately 6 minutes. After the exposure block and a brief pause to collect the pen and paper from the participant, the movie and stimuli presentation were started again. The experiment lasted approximately 45 minutes including breaks, and was presented using E-Prime 2.0.10 (Psychology Software Tools, Sharpsburg, Pennsylvania, USA). The whole session, including preparations, lasted about an hour and a half.

2.5. EEG

The BioSemi ActiveTwo system with ActiView software was used for electroencephalography (EEG) data recording (BioSemi, Amsterdam, The Netherlands). A driven-leg reference was used (CMS/DRL loop with voltage recorded relative to the CMS electrode), and the sampling rate was 2048 Hz. Sixteen head electrodes were used (Fp1, Fp2, F3, F4, Fz, T7, T8, C3, C4, Cz, P3, P4, Pz, O1, O2 and Oz). Electrodes were also placed above and below the left eye and outside the lateral canthus of each eye (for automatic identification of eye-movements), as well as behind each ear on the mastoid bones (for use as reference channels). Preprocessing of data was performed as described in [16], except that the independent component analysis was performed on epoched rather than continuous data. The MMN amplitude was calculated as the mean amplitude of channel Fz in the time window 150-300 ms in the subject average difference waveform (subject average standard waveform subtracted from subject average deviant waveform). Time window and channel were chosen based on where a strong MMN response is typically found [19]. Statistical analyses were performed in R [22].

3. Results

To test whether the current design evoked an MMN response in the participants, a one-sample t-test was performed on the MMN amplitude in the pre-exposure block. The reason for using only the pre-exposure block for the MMN check is that the absence, existence and/or amplitude of an MMN in the post-exposure could be impacted by the experimental manipulation (i.e., the exposure). The t-test revealed that an MMN was indeed elicited (t(15) = -2.8489, p = .006). Figure 2 shows the mean MMN amplitude for the unimodal and the bimodal groups in the pre-exposure and the post-exposure blocks, and Figure 3 shows the grand average waveforms for standard and deviant at Fz, as well as the difference waveforms.

In order to test the effect of exposure to the different distributions on the MMN amplitude, a 2x2 repeated measures ANOVA was carried out. Within-subject variable was Block (pre-exposure vs. post-exposure) and between-subjects variable was Distribution (unimodal vs. bimodal).

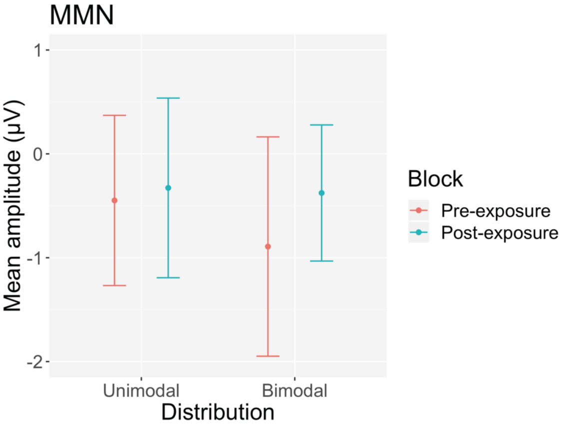

Figure 2: *Mean MMN amplitudes for pre- and post-exposure blocks for participants exposed to the unimodal and the bimodal distribution respectively. Error bars show the standard deviation.*

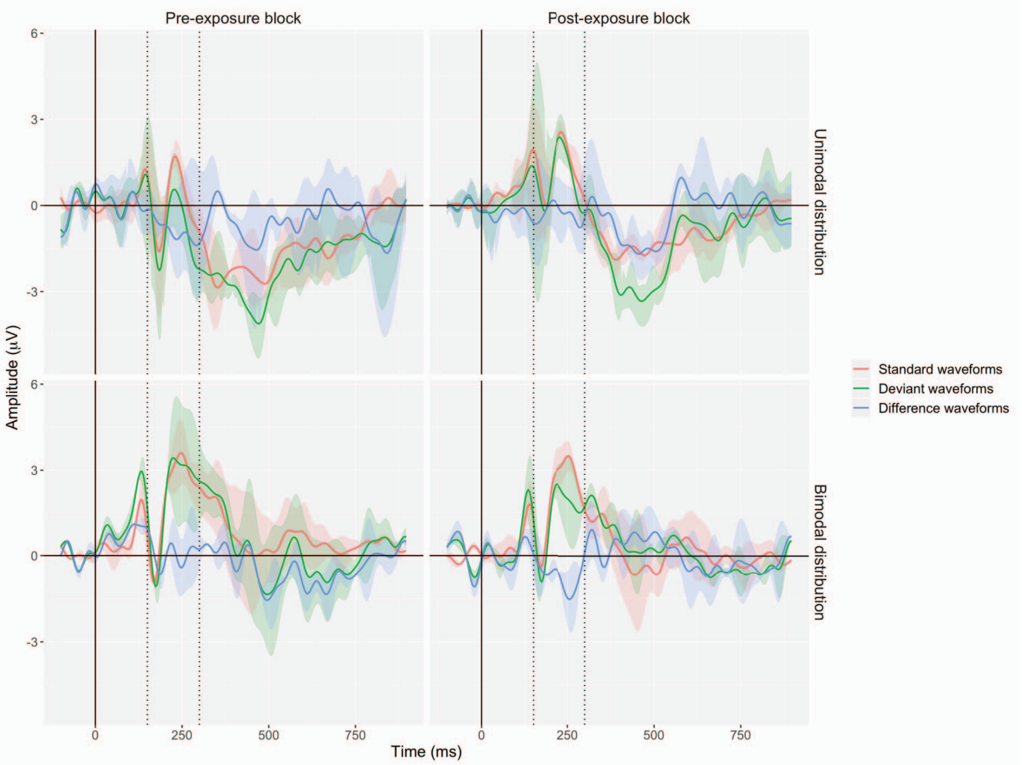

Figure 3: *The grand-average for standard (red), deviant (green) and difference (blue) waveforms obtained at the Fz electrode, for the unimodal group (top) and the bimodal group (bottom), and for the pre-exposure block (left) and the post-exposure block (right). The analysis time window is denoted by vertical dotted lines.*

The main effect of Distribution was not significant (F(1,14) = 1.067, *p* = .312), that is, there was no overall difference between participant groups when blocks were collapsed. The main effect of Block was not significant (F(1,14) = 0.263, *p* = .613), which means that there was no difference in amplitude between the pre-exposure block and the post-exposure block when both groups were collapsed. The interaction between Block and Distribution was not significant (F(1,14) = 2.318, *p* = .141), that is, the relative MMN amplitude across blocks did not differ between the two participant groups.

4. Discussion

In the present study, no evidence of DL was found for attended listening to complex non-speech in adults[1]. The lack of DL reported in our previous study [16] is thus presumably not due to unattended listening to the stimuli during exposure. This pattern is not in line with previous research, which has shown an effect of attention in adults [15], and in 10-month-old infants [17]. However, these studies used speech stimuli, which means that DL shifted boundaries in already existing perceptual categories, whereas our studies attempted to demonstrate DL for stimuli with no pre-existing categories.

One possible explanation for the lack of DL for non-speech, even under attended listening conditions, is that it simply does not occur, either because the process only operates on sounds recognized as species-specific, such as vocalizations, or because it operates on pre-existing perceptual categories only. An alternative explanation is that new categories start to form, but that the MMN is not sensitive enough to capture these first signs of perceptual categorization.

Auditory perceptual categorization has been demonstrated for complex non-speech stimuli [23], [24], but this involved active (although implicit) training and the presence of invariant visual category cues. Perceptual non-speech categories activate the brain areas tied to speech sound processing [25], suggesting that they can be formed under the same conditions as speech sound categories. Spectrally rotated speech specifically, can be understood after explicit training [20]. This argues against the notion that DL does not occur for sounds that are not species-specific, provided that it is in fact involved in the formation of new perceptual categories, such as in the very first stages of speech sound category development.

It is however possible that DL is not involved in forming novel perceptual categories, at least not under the conditions of a typical DL experiment. In all DL studies to date, both with adults (*e.g.*, [12], [13]) and infants (*e.g.*, [4], [5]), DL has been shown to modify already existing categories. To test whether pre-existing categories are needed for the DL process to operate, perceptual non-speech categories can be induced through implicit or explicit training (*e.g.*, [23], [24], [26]) after which DL can be tested on those categories).

---

[1] In the present study, participants listened to 340 tokens during exposure. This is a substantially larger number than has been used in previous studies where DL has been found in adults [28], [29], so the lack of DL is likely not due to too little exposure.

Finally, although the MMN has been used both to detect DL in infants [10] and to assess perceptual categorization of non-speech induced under other conditions in adults [18], it is possible that it does not capture the very first signs of perceptual categories, or that the MMN is not as readily modulated in adults as in infants [27]. To test whether DL of non-speech can be evidenced in adults by other measures, a study using a behavioral paradigm (similar to [12], [13]) is currently underway.

In conclusion, no evidence of DL has been demonstrated for complex non-speech in adults, neither when participants listen attentively or when they do not attend to stimuli during exposure [16]. There are several possible explanations for this. The ones deemed most likely are that DL does not operate on stimuli for which there are no pre-existing perceptual categories, or that the MMN does not capture the very first signs of perceptual category formation. Studies are underway to rule out methodological questions.

5. Author contributions

Study conceptualization: EM; Experiment design: EEC, EM, JS; Experiment and stimuli creation: EEC, EM, JS; Data collection: EM (incl. supervision of students), LG; Data processing: EM; Analysis and interpretation: EEC, EM, JS, LG; Drafting of the manuscript: EM; Critical revisions of the manuscript: EEC, EM, JS, LG.

6. Acknowledgements

The study was funded by Riksbankens Jubileumsfond (RJ) through the project “Effects of hyperarticulation on early language development (HELD)” (RJ P17-0175:1, PI Gustavsson). The authors would like to thank all participants for their contributions, the students Johanna Paavig, Mary Martinsson and Maryan Muse Mohamed for help with data collection, and Klara Hjerpe for help with data collection.

7. References

[1] B. McMurray, R. Aslin and J. Toscano, "Statistical learning of phonetic categories: insights from a computational approach," *Developmental Science,* vol. 12, pp. 369-378, 2009.

[2] D. Kleinschmidt and T.F. Jaeger, "A Bayesian Belief Updating Model of Phonetic Recalibration and Selective Adaptation," in *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, Portland, Oregon, 2011.

[3] D. Kleinsmidt and T.F. Jaeger, "Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning?," *Psychonomic bulletin & review,* vol. 23, no. 3, pp. 678-691, 2016.

[4] J. Maye, J.F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition,* vol. 82, no. 3, p. B101–B111, 2002.

[5] J. Maye, D.J. Weiss, and R.N. Aslin, "Statistical phonetic learning in infants: Facilitation and feature generalization," *Developmental science,* vol. 11, no. 1, p. 122–134, 2008.

[6] K.A. Yoshida, F. Pons, J. Maye, and J.F. Werker, "Distributional phonetic learning at 10 months of age," *Infancy,* vol. 15, no. 4, pp. 420-433, 2010.

[7] A. Cristià., G.L. McGuire, A. Seidl, and A.L. Francis, "Effects of the distribution of acoustic cues on infants' perception of sibilants," *Journal of phonetics,* vol. 39, no. 3, pp. 388-402, 2011.

[8] L. Liu and R. Kager, "How do statistical learning and perceptual reorganization alter Dutch infant’s perception to lexical tones?," in *17th International Congress of Phonetic Sciences*, Hong Kong, 2011.

[9] L. Liu and R. Kager, "Statistical learning of speech sounds is most robust during the period of perceptual attunement," *Journal of Experimental Child Psychology,* vol. 164, pp. 192-208, 2017.

[10] K. Wanrooij, P. Boersma., and T. Van Zuijen, "Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study," *Frontiers in psychology,* vol. 5, no. 77, 2014.

[11] P. Escudero, T. Benders, and K. Wanrooij, "Enhanced bimodal distributions facilitate the learning of second language vowels," *The Journal of the Acoustical Society of America,* vol. 130, no. 4, pp. EL206-EL212, 2011.

[12] J. Maye and L. Gerken, "Learning phonemes without minimal pairs," in *Proceedings of the 24th annual Boston University Conference on Language Development, vol. 2*, Boston, 2000.

[13] J. Maye and L. Gerken, "How far can the input take us?," in *Proceedings of the 25th annual Boston University Conference on Language Development, vol. 1*, Boston, 2001.

[14] P. Escudero and D.Williams, "Distributional learning has immediate and long-lasting effects," *Cognition,* vol. 133, no. 2, pp. 408-413, 2014.

[15] J.H. Ong, D. Burnham and P. Escudero, "Distributional learning of lexical tones: A comparison of attended vs. unattended listening," *PloS one,* vol. 10, no. 7, p. e0133446, 2015.

[16] E. Marklund, E.E. Cortes, and J. Sjons, "MMN responses in adults after exposure to bimodal and unimodal frequency," in *Interspeech 2017*, Stockholm, 2017.

[17] K. Yoshida, "Plasticity in infants' speech perception: a role for attention?," *Ph.D. dissertation, University of British Columbia,* 2008.

[18] R. Liu and L.L. Holt, "Neural Changes Associated with Nonspeech Auditory Category Learning Parallel Those of Speech Category Acquisition," *Journal of Cognitive Neuroscience,* vol. 23, no. 3, 2011.

[19] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, "The mismatch negativity (MMN) in basic research of central auditory processing: A review," *Clinical Neurophysiology,* vol. 118, no. 12, pp. 2544-2590, 2007.

[20] B. Blesser, "Speech Perception Under Conditions of Spectral Transformation: I. Phonetic Characteristics," *Journal of Speech and Hearing Research,* vol. 15, no. 1, pp. 5-41, 1972.

[21] E. Marklund, F. Lacerda, and I-C. Schwarz, "Using rotated speech to approximate the acoustic mismatch negativity response to speech," *Brain and Language,* vol. 176, pp. 26-35, 2018.

[22] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2018.

[23] Y. Gabay and LL. Holt, "Incidental learning of sound categories is impaired in developmental dyslexia," *Cortex,* vol. 73, pp. 131-143, 2015.

[24] T. Wade and L.L. Holt, "Effects of later-occurring nonlinguistic sounds on speech categorization," *The Journal of the Acoustical Society of America,* vol. 118, no. 3, pp. 1701-1710, 2005.

[25] R. Leech, L.L. Holt, J.T. Devlin, and F. Dick, "Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions," *Journal of Neuroscience,* vol. 29, no. 16, pp. 5234-5239, 2009.

[26] L.L. Holt and A.J. Lotto, "Cue weighting in auditory categorization: Implications for first and second language acquisition," *The Journal of the Acoustical Society of America,* vol. 190, no. 5, 2006.

[27] K. Wanrooij, P. Boersma, and T.L. van Zuijen, "Distributional vowel training is less effective for adults than for infants. A study using the mismatch response," *PloS one,* vol. 9, no. 10, p. e109806, 2014.

[28] K. Wanrooij, P. Escudero, and M. E. Raijmakers, "What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning," *Journal of Phonetics,* vol. 41, no. 5, pp. 307-319, 2013.

[29] J. Gilkerson, "Categorical perception of natural and unnatural categories: Evidence for innate category boundaries," *UCLA working papers in linguistics,* vol. 13, pp. 34-58, 2005.