Effect of rhythmic attention on the segregation of interleaved melodies

Aymeric Devergie, Nicolas Grimault, Barbara Tillmann, Frédéric Berthommier

Effect of rhythmic attention on the segregation of interleaved melodies

Aymeric Devergie, Nicolas Grimault [a)], and Barbara Tillmann

Laboratoire de Neurosciences Sensorielles,

Comportement et Cognition,

CNRS UMR 5020,

Université Lyon 1,

Lyon,

FRANCE

aymeric.devergie@olfac.univ-lyon1.fr,

nicolas.grimault@olfac.univ-lyon1.fr,

barbara.tillmann@olfac.univ-lyon1.fr

Frédéric Berthommier

GIPSA-Lab,

CNRS UMR 5216,

Université Stendhal,

Grenoble,

FRANCE

Frederic.Berthommier@gipsa-lab.grenoble-inp.fr

(Dated: May 4, 2010)

[a)] author to whom correspondence should be addressed

Abstract

As previously suggested, attention may increase segregation via enhancement and suppression sensory mechanisms. To test this hypothesis, we proposed an interleaved melody paradigm with two rhythm conditions applied to familiar target melodies and unfamiliar distractor melodies sharing pitch and timbre properties. When rhythms of both target and distractor were irregular, target melodies were identified above chance level. A sensory enhancement mechanism guided by listeners knowledge may have helped to extract targets from the interleaved sequence. When the distractor was rhythmically regular, performance was increased, suggesting that the distractor may have been suppressed by a sensory suppression mechanism.

PACS numbers: 43.66 Mk, 43.66 Ba

Keywords: rhythmic attention, interleaved melodies, auditory scene analysis, knowledge

I. INTRODUCTION

In everyday listening, sound events rarely appear in isolation. Usually, several acoustic streams issued from various sources compete with each other. According to Bregman (1990), segregation mechanisms based on listeners' knowledge and schemata stored in long-term memory can be used to extract a well-known target from an acoustic mixture. It is generally acknowledged that these extraction mechanisms are strongly related to attentional processes (Hafter *et al.*, 2003). Our present study is aimed at further elucidating the attentional involvement in these extraction mechanisms, referred to fission mechanisms (Moore and Gockel, 2002), for sequential stream segregation.

Numerous studies have implicated differences in pitch, loudness or timbre as mediating the low-level processing of segregation. In contrast, few studies have investigated schema-based processing per se. Focusing on fission mechanisms, Dowling *et al.* (1987) proposed an interleaved melody paradigm to test whether listeners are able to identify familiar melodies (targets) interleaved with unknown melodies (distractors). In one condition in which target and distractor melodies shared the same pitch range and timbre, listeners were still able to extract the target melody from the interleaved sequences. The authors concluded that listeners performed the fission task by using their prior knowledge of the target melody. In fact, in this study, participants knew which melody to listen to and to extract from the interleaved sequences. In this situation, the involved attentional processes are guided by familiar melodic schemata stored in memory, and these schemata likely help to extract the relevant information from an auditory mixture.

In the study by Dowling *et al.*, the interleaved sequences (i.e. target + distractor) started either with a tone from the target melody (on-beat condition) or a tone from the distractor melody (off-beat condition). Target identification was found to be 10% better in the on-beat condition than in the off-beat condition. This result is consistent with Jones' theory of rhythmic attention, which is described as a dynamic process that builds up temporal windows of expectation (Jones, 1976; Jones *et al.*, 1981, 2002). For example, Jones *et al.* (2002) showed that comparing the pitch of two tones was easier when the second tone occurred within an expected time window and was primed by a rhythmic 8-tone sequence. Furthermore, Dowling *et al.* (1987) showed that attentional processes were primed by starting the sequence with a tone of the target melody.

Rhythmic attention may have also contributed to performance in streaming studies testing the effect of the rhythm of presentation on sequential segregation. Van Noorden (1975) and George and Bregman (1989) addressed this idea using an experimental setup that biased listeners toward perceiving a single stream. In the Van Noorden (1975) study, listeners were asked to adjust the pitch difference of tones from two streams in order to perceive all tones integrated into a single stream. Listeners' adjustments were not found to be influenced by the regular or random nature of the tone rhythm. In the George and Bregman (1989) study, listeners were able to integrate two tone sequences into one single stream regardless of the tone rhythms. These findings suggest that the ability to integrate all events into one stream, reputed to be an automatic (bottom-up) process (Bregman, 1990), is independent of rhythmic attention. In contrast, the only study that investigated the effect of rhythm on the ability to selectively listen to part of an acoustic mixture (fission), reported a strong effect of rhythm (Jones *et al.*, 1981). This finding is consistent with Van Noorden (1975) who reported that fission can be influenced by top-down processes.

Using a similar experimental setup as Bregman and Rudnicky (1975), Jones *et al.* (1981) asked participants to judge the temporal order of two tones that were embedded in a sequence of captor tones with lower pitch. They found that performance was weaker when all tones were played with an isochronous rhythm at the same speed than when the probe tones and the captor tones were played at different rates of speed. Furthermore, the segregation of the probe tones induced by the pitch difference was enhanced by the tempo difference. The authors interpreted these results in terms of rhythmic attention.

In addition, for primitive, stimulus-driven segregation (Bregman, 1990), the effect of attention remains a matter of debate. Presuming that attention might be an all or none process, some authors have argued that primitive segregation can be influenced by attention

(Carlyon *et al.*, 2001), while others have argued that it is purely pre-attentive (Sussman *et al.*, 2007). Two recent studies suggested that primitive segregation can be decomposed into pre-attentive and attentive mechanisms (Snyder and Alain, 2007; Cusack *et al.*, 2004). These attentional mechanisms, if they exist, would interact with stimulus-driven segregation and would be related to acoustic cues. For schema-based segregation, the effect of attention has generally been acknowledged (Bregman, 1990). However, while top-down attention could focus on a limited range of acoustic features (Hafter *et al.*, 2003), there are only few experimental data showing the effect of attention on schema-based segregation (Dowling *et al.*, 1987).

Despite the lack of experimental data, two theoretical frameworks on attention and segregation have recently been proposed (Fritz *et al.*, 2007; Alain and Bernstein, 2008). Fritz *et al.* (2007) suggested that two attentional mechanisms might be involved in auditory segregation, a bottom-up 'pop-out' process and top-down mechanism. In the bottom-up 'pop-out' process, acoustic features (i.e., pitch, timbre or rhythmic regularity) catch listeners' attention and enhance the processing of the relevant acoustic signals. The top-down mechanism is based on the development of expectancies derived from listeners' knowledge. Alain and Bernstein (2008) suggested a complementary theoretical background in which attention increases segregation via two mechanisms. The first mechanism enhances the processing of task-relevant material, and the second mechanism, a suppression mechanism, attenuates the processing of task-irrelevant material.

Our experiment was conducted to provide some new data to test hypothesis derived from the frameworks of Fritz *et al.* (2007); Alain and Bernstein (2008). Listeners were required to extract a relevant target interleaved with a distractor. The only available cues for segregation were the knowledge about the target and, in some condition, the regularity of the rhythm of part of the signal, which needed to be ignored to test for suppression.

II. EXPERIMENT

A. Rationale

In the current study, an interleaved melody task was designed in which listeners were instructed to identify familiar target melodies embedded in distractor melodies sharing the same pitch and timbre ranges . Thus, neither pitch cues (Dowling *et al.*, 1987) nor timbre cues (Bey, 1999) would be useful for segregation in this task. Two conditions were utilized. In condition 1, the rhythms of target and distractor melodies were irregular. In condition 2, the rhythm of the target melody was irregular, while the rhythm of the distractor melody was regular. The rationale for using two conditions was two-fold. First, in condition 1, identification performance at above chance levels would indicate an enhancement mechanisms based on knowledge. Second, in condition 2, any increase of performance relative to condition 1 would indicate a suppression mechanism that is strengthened by the regular rhythm of the distractor melody.

B. Apparatus

1. Participants

Twenty participants aged 18-30 years (mean=22.7, s.d=2.1) participated in the experiment. All participants were native French speakers and had pure tone audiometric thresholds below 15 dB HL at octave frequencies between 250 and 4000 Hz (American National Standards Institute, 2004). All participants were paid an hourly wage for their participation and signed an informed consent form. This study was formally approved by a local ethics committee (CPP Sud-Est II No. 06035).

2. Stimuli

Eight familiar French *target* melodies (displayed in Figure 1) were selected and rendered isochronous. In addition, corresponding *control* melodies matched to each familiar target melody were constructed to analyze listeners' potential use of pitch range cues (see below). The control melodies were constructed by randomly selecting one temporal order of the notes of the familiar target melody among all order possibilities avoiding note repetition. In the interleaved melody task, each of the target and control melodies was mixed with a corresponding *distractor* melody (Figure 2). The pitches of the distractor notes were chosen to be within the pitch range corresponding to the maximal pitch range across the eight familiar melodies (i.e., between 196 Hz and 392 Hz, g2 and g3 in the musical scale). The target and control melodies were interleaved note-by-note with distractor melodies. Pitches of the distractor notes were randomly chosen without repetition of successive notes.

All notes of all melodies lasted 80 ms, including 10 ms rising and falling ramps. Each of these notes was created with five French vowels /a/, /i/, /e/, /o/, /y/ with various pitches (Figure 1). The vowels were generated using the Klatt algorithm (Klatt, 1980), and the specific spectral content of each vowel (i.e., formant positions) introduced some timbre variations across vowels within a sequence. As indicated by Singh and Bregman (1997), such timbre variation reduces the global perceptual coherence of the sequence and may contribute to segregation. Vowels of the interleaved sequences were randomly chosen without direct repetition of vowels for successive notes.

[FIG. 1 about here.]

Because each vowel was used for target, control and distractor melodies, timbre was not a reliable cue for the segregation and identification task. Within each sequence, the pitch range of the target melody fell within the pitch range of the distractor melody. Moreover, each target melody shared exactly the same pitch range with the corresponding control melodies. To test for the listeners' use of pitch range cues in the segregation task, which would predict identification of a control melody as the associated target melody, we directly compared performance for control and target melodies.

Two different rhythm conditions were defined (Figure 2). The left panel of Figure 2 represents condition 1, in which notes of target or control melodies and notes of distractor melodies had random onsets. In this condition, the inter-onset-interval (IOI) separating two successive notes (first-order interval) was randomly chosen between 80 and 140 ms. The right panel of Figure 2 represents condition 2, in which notes from the target or control melodies had random onsets, and notes from the distractor melodies had regular onsets. Temporal regularity, in the form of isochrony, was introduced by setting all time intervals separating the onsets of two successive distractor notes (second-order intervals) equal to 300 ms. As in condition 1, IOIs between a target note (or a control note) and a distractor note were randomly chosen between 80 and 140 ms. A Chi-square analysis performed on the IOI distributions revealed that they were not significantly different [$\chi^2 = 19.81$, $p = 0.997$] for these two conditions. The target melodies were the same for the two rhythm conditions.

[FIG. 2 about here.]

Each of the 8 familiar melodies was repeated four times in association with a different distractor melody. Overall, 32 different distractors were generated, and the same distractor melodies were combined with the control and target melodies. For example, target melody 1 was associated to the same four distractors as control melody 1. The 64 resulting combinations were then duplicated for the two rhythm conditions, yielding 128 experimental trials. Thus, the same interleaved sequences were used in the two rhythm conditions. All interleaved sequences started with a distractor note, and not a target note (leading to the target melody being an off-beat melody, as in Dowling *et al.* (1987)). This was done to avoid that the first tone of the target melody might attract attention and help listeners to perform the task,

The only difference between the two conditions was the rhythm of the distractor stimuli played using a SIGMATEL internal sound card connected to a Sennheiser HD 250 Linear II headphone. Listeners were comfortably seated in a double-walled attenuated sound booth. Output level was calibrated to 70 dB SPL (Larson Davis AEC101 and 824; American National Standards Institute (1995)) with RMS value adjustment between all vowels.

3. Procedure

Before the main experimental task (i.e., identification in the interleaved melodies), listeners performed first a familiarization task, then an identification task on the target melodies. In these two preliminary tasks, the target melodies were generated alone with an isochronous rhythm (IOI of 384 ms). In the familiarization task, the titles of the eight melodies were displayed and the listeners was instructed to listen to each melody as many times they want by selecting the corresponding title. In the initial identification task, the participants listened to the melodies in a random order and were instructed to identify the target melody as fast as possible. The titles of the eight melodies were displayed. An additional title 'unknown melody' was also displayed. Two participants did not reach 50% correct identification and were excluded from further testing. The identification performance averaged across the remaining participants is indicated for each melody in Figure 1. A high percent of correct target melody identification was reached despite rhythms being isochronous. This finding is consistent with Hébert and Peretz (1997), who reported that pitch contour is a dominant feature used for identification of familiar melodies. This also suggests that identification should remain high in the interleaved melody task even if the rhythm of the target melodies was rendered irregular. In the interleaved melody task, all sequences were played in random order for each participant. Participants had to identify the target melody present in the sequence.

C. Results

Responses for control melodies were averaged for each participant. The results showed that 65% (s.d=5.2) of the control melodies were categorized as unknown melodies, 5% (s.d=2.4) were identified as the associated target melody and 4% (s.d=2) were identified as a different target melody than the associated target. A t-test applied to the two latter identification scores did not reveal a significant bias toward the associated target melody [$t(62)=1.29$; $p=0.2$]. These findings indicate that the pitch range of the familiar melody was not used as a reliable cue allowing melody identification. Therefore, neither pitch range nor timbre range aided in segregating target melodies from distractor melodies in our study.

Responses for target melodies were considered correct when the target melody was correctly identified. Figure 3 shows a plot of identification performance for each rhythm condition averaged over all participants. A t-test revealed that performance was significantly better than chance in condition 1 [$t(17)$=2.87; p=0.01] and in condition 2 [$t(17)$=2.62; p=0.018]. In addition, performance was significantly better in condition 2 than in condition 1 [$t(17)$=2.28; p=0.036], indicating an effect of rhythm presentation.

In condition 1, the only cue available for participants was their prior knowledge of the familiar melodies. To test whether listeners' prior knowledge could explain performance in the interleaved melody task, we evaluated the correlation between identification scores of individual melodies in the familiarization phase and identification scores in the interleaved melodies task. We found that better identification of the familiar melody presented alone tended to be correlated with better identification of the target melody in the interleaved sequences [R^2=0.4655, p=0.0623]. Due to the small number of data points available, an additional *Monte-Carlo* simulation analysis was also applied to the correlation data. This analysis yielded a similar significance value [p=0.055].

[FIG. 3 about here.]

III. DISCUSSION

Familiar melody identification in the interleaved melody task was above chance regardless of the rhythm of the distractor and was better when the rhythm of the distractor was regular. In retrospect, these performance levels suggest that both the stimuli and the task were appropriate to test our hypotheses.

Condition 1 was designed to test the theory proposed by Alain and Bernstein of the existence of an enhancement mechanism based on listeners' knowledge, involved in auditory fission. All previous studies measuring fission boundary used stimuli with either acoustic cues (Jones *et al.*, 1981) (for a review, see Fritz *et al.* (2007)) or rhythm cues (Dowling *et al.*, 1987). The better than chance performance in condition 1 provides the first evidence for a pure attentional fission mechanism based only on prior knowledge of a pitch sequence (decoupled from its original rhythm). In fact, listeners' knowledge was the only reliable cue that enabled identification of target melodies. Thus, segregation can be influenced by knowledge when the schemata are stored in long term memory (Dowling *et al.* (1987), this study). In contrast, by using unfamiliar melodies presented just before the test to produce schemata stored only in short term memory, Bey and McAdams (2002) reported performances at chance levels when acoustical cues were lacking. These apparently divergent findings may be reconciled by the strength of knowledge being important for segregation, as previously hypothesized by Bey (1999). Our results are consistent with this hypothesis (Bey, 1999) as indicated by the positive correlation tendency between identification scores in the familiarization phase and identification in condition 1. The melodies were reputed to be familiar to native French speakers and, thus, more likely to be stored in long term memory.

In addition to the influence of knowledge stored in long term memory, Dowling *et al.* (1987) showed the influence of the rhythmic position of the target melody (i.e. on- or off-beat). The improved performance during the on-beat condition was consistent with rhythmic attention described by Jones (1976), as discussed in the Introduction section. Although our target melodies were off-beat, our findings are consistent with the rhythmic attention theory. Jones *et al.* (2002) provided information regarding the size of the attentional window. Based on their data (Figure 4 from Jones *et al.* (2002)), the size of the expectancy window, defined by a performance decrease of 10%, was a few tens of milliseconds. In our experiment, the rhythm of the target melodies was always irregular, but at least part of each target note fell within the expectancy window. Despite randomized IOIs and the hypothesis that expectancy windows could vary with contextual irregularity, the results of the current experiment were still consistent with the rhythmic attention theory.

Comparison of conditions 1 and 2 allowed testing for the effect of rhythmic attention on the suppression mechanism. Our data revealed that identification increased significantly when the rhythm of the distractor was regular. This finding is consistent with those of Demany and Semal (2002), who showed that second-order temporal regularity could be beneficial for perception. In our experiment, this second-order regularity presumably helped to build a rhythm attention cycle (Jones *et al.*, 2002) synchronized to the notes of the distractor melody.

In sum, our current study provides the first evidence of pure attentional segregation based on knowledge that can be strengthened by rhythm regularity of the part of the signal that needs to be suppressed. These results are consistent with the rhythmic attention theory of Jones and Boltz (1989) and demonstrate the relevance of rhythmic attention for auditory scene analysis.

Acknowledgments

This work was supported by grants from the Région Rhones-Alpes Auvergne 'Cluster HVN 2007' and the Agence Nationale de Recherche (ANR-08-BLAN-0167-01). Special thanks to Mary Riess Jones for very helpful comments on a previous version of the manuscript, to Jay Dowling for interesting suggestions and to Charlotte Dépalle for managing participants.

References

Alain, C. and Bernstein, L. J. (**2008**). "From sounds to meaning: the role of attention during auditory scene analysis.", Curr Opin Otolaryngol Head Neck Surg **16**, 485–489.

American National Standards Institute (**1995**). "Ansi s3.7-r2003: Methods for coupler calibration of earphones", .

American National Standards Institute (**2004**). "Ansi s3.21-2004: Methods for manual pure-tone threshold audiometry", .

Bey, C. (**1999**). "Reconnaissance de mélodies intercalées et formation des flux auditifs: Analyse fonctionnelle et exploration neuropsychologique", Unpublished phdthesis.

Bey, C. and McAdams, S. (**2002**). "Schema-based processing in auditory scene analysis.", Percept Psychophys **64**, 844–854.

Bregman, A. (**1990**). *Auditory Scene Analysis: The Perceptual Organzation of Sounds* (MIT Press).

Bregman, A. S. and Rudnicky, A. I. (**1975**). "Auditory segregation: stream or streams?", J Exp Psychol Hum Percept Perform **1**, 263–267.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). "Effects of attention and unilateral neglect on auditory stream segregation.", J Exp Psychol Hum Percept Perform **27**, 115–127.

Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (**2004**). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis.", J Exp Psychol Hum Percept Perform **30**, 643–656.

Demany, L. and Semal, C. (**2002**). "Limits of rhythm perception.", Q J Exp Psychol A **55**, 643–657.

Dowling, W. J., Lung, K. M., and Herrbold, S. (**1987**). "Aiming attention in pitch and time in the perception of interleaved melodies.", Percept Psychophys **41**, 642–656.

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (**2007**). "Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1?", Hear Res **229**, 186–203.

George, M. F.-S. and Bregman, A. S. (**1989**). "Role of predictability of sequence in auditory stream segregation.", Percept Psychophys **46**, 384–386.

Hafter, E. R., sarampalis, A., and Psyche, L. (**2003**). *Springer handbook of auditory research: Auditory perception of sound sources*, volume 5, chapter Auditory Attention and Filters, 115–142 (Springer).

Hébert, S. and Peretz, I. (**1997**). "Recognition of music in long-term memory: are melodic and temporal patterns equal partners?", Mem Cognit **25**, 518–533.

Jones, M., Moynihan, H., MacKenzie, N., and Puente, J. (**2002**). “Temporal aspects of stimulus-driven attentding in dynamic arrays”, Psychological Science **13**, 313–319.

Jones, M. R. (**1976**). “Time, our lost dimension - toward a new theory of perception, attention and memory”, Psychological Review **83**, 323–355.

Jones, M. R. and Boltz, M. (**1989**). “Dynamic attending and responses to time.”, Psychol Rev **96**, 459–491.

Jones, M. R., Kidd, G., and Wetzel, R. (**1981**). “Evidence for rhythmic attention.”, J Exp Psychol Hum Percept Perform **7**, 1059–1073.

Klatt, D. (**1980**). “Software for a cascade/parallel formant synthesizer”, J Acoust Soc Am **67**, 971–995.

Moore, B. C. J. and Gockel, H. (**2002**). “Factors influencing sequential stream segregation”, Acta Acustica **88**, 320–333.

Singh, P. G. and Bregman, A. S. (**1997**). “The influence of different timbre attributes on the perceptual segregation of complex-tone sequences.”, J Acoust Soc Am **102**, 1943–1952.

Snyder, J. S. and Alain, C. (**2007**). “Toward a neurophysiological theory of auditory stream segregation.”, Psychol Bull **133**, 780–799.

Sussman, E. S., Horvath, J., Winkler, I., and Orr, M. (**2007**). “The role of attention in the formation of auditory streams.”, Percept Psychophys **69**, 136–152.

Van Noorden, L. (**1975**). “Temporal coherence in the perception of tone sequences”, Unpublished doctoral dissertation, Technische Hogeschool Eindhovern, Eindhoven, The Netherlands.

List of Figures

List of Figures

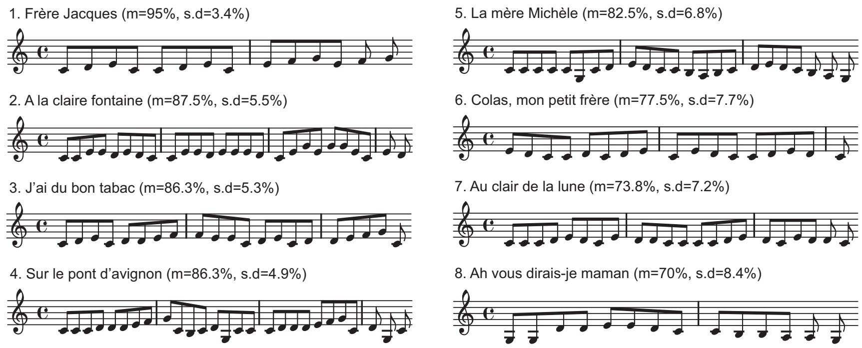


FIG. 1. Musical scores of the eight familiar French melodies. The results from the identification task where the melodies were presented alone are indicated in brackets (mean percent of correct identification, standard deviation)

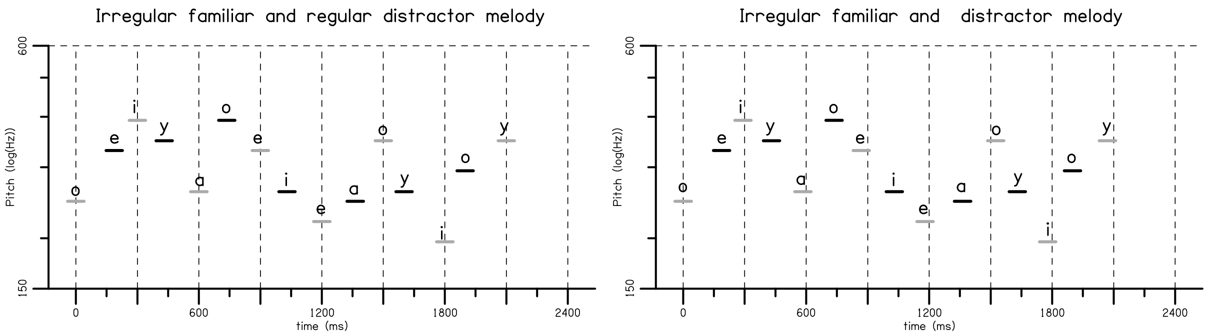


FIG. 2. Schematic representation of an interleaved sequence in the two rhythm conditions. An excerpt from the familiar melody 'Sur le pont d'Avignon' is represented by the black lines and one of the 32 possible distractor melodies is represented by the gray lines.

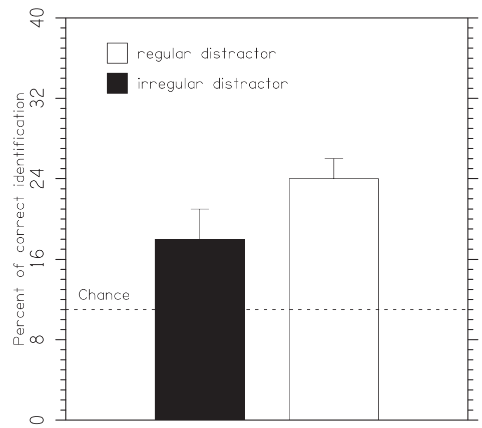


FIG. 3. Percent of correctly identified target melodies interleaved with irregular distractors (black bar) or regular distractors (white bar). Chance level is equal to 11% (1/9 possibilities). Error bars represent standard deviation.

1. Frère Jacques (m=95%, s.d=3.4%)

2. A la claire fontaine (m=87.5%, s.d=5.5%)

3. J’ai du bon tabac (m=86.3%, s.d=5.3%)

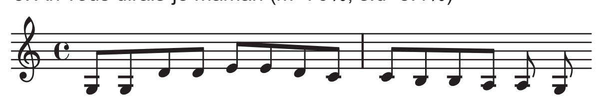
4. Sur le pont d’avignon (m=86.3%, s.d=4.9%)

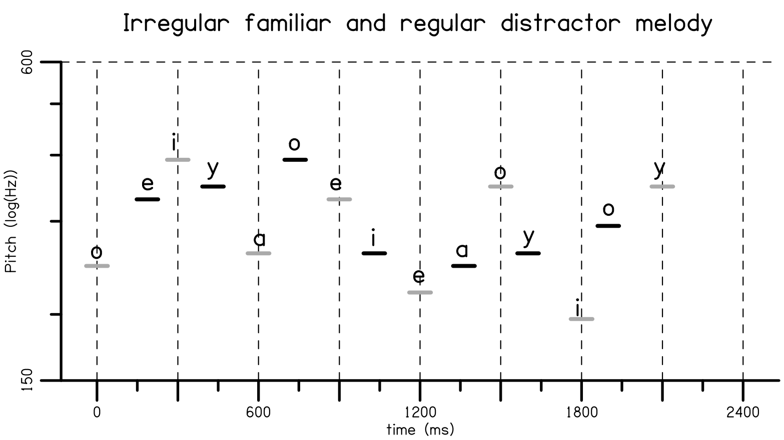
5. La mère Michèle (m=82.5%, s.d=6.8%)

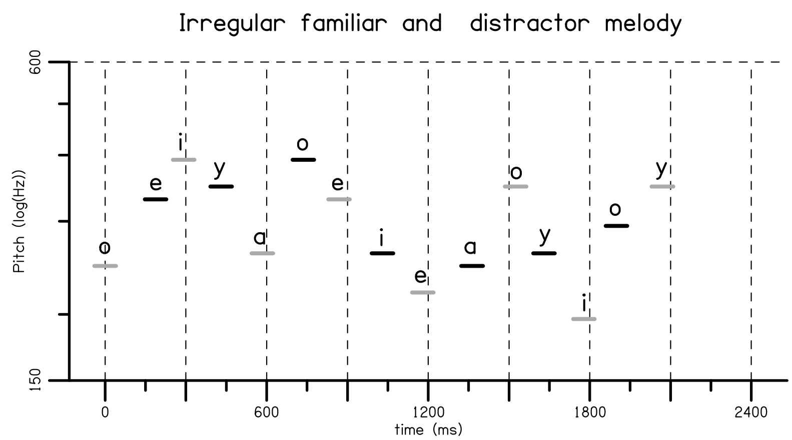
6. Colas, mon petit frère (m=77.5%, s.d=7.7%)

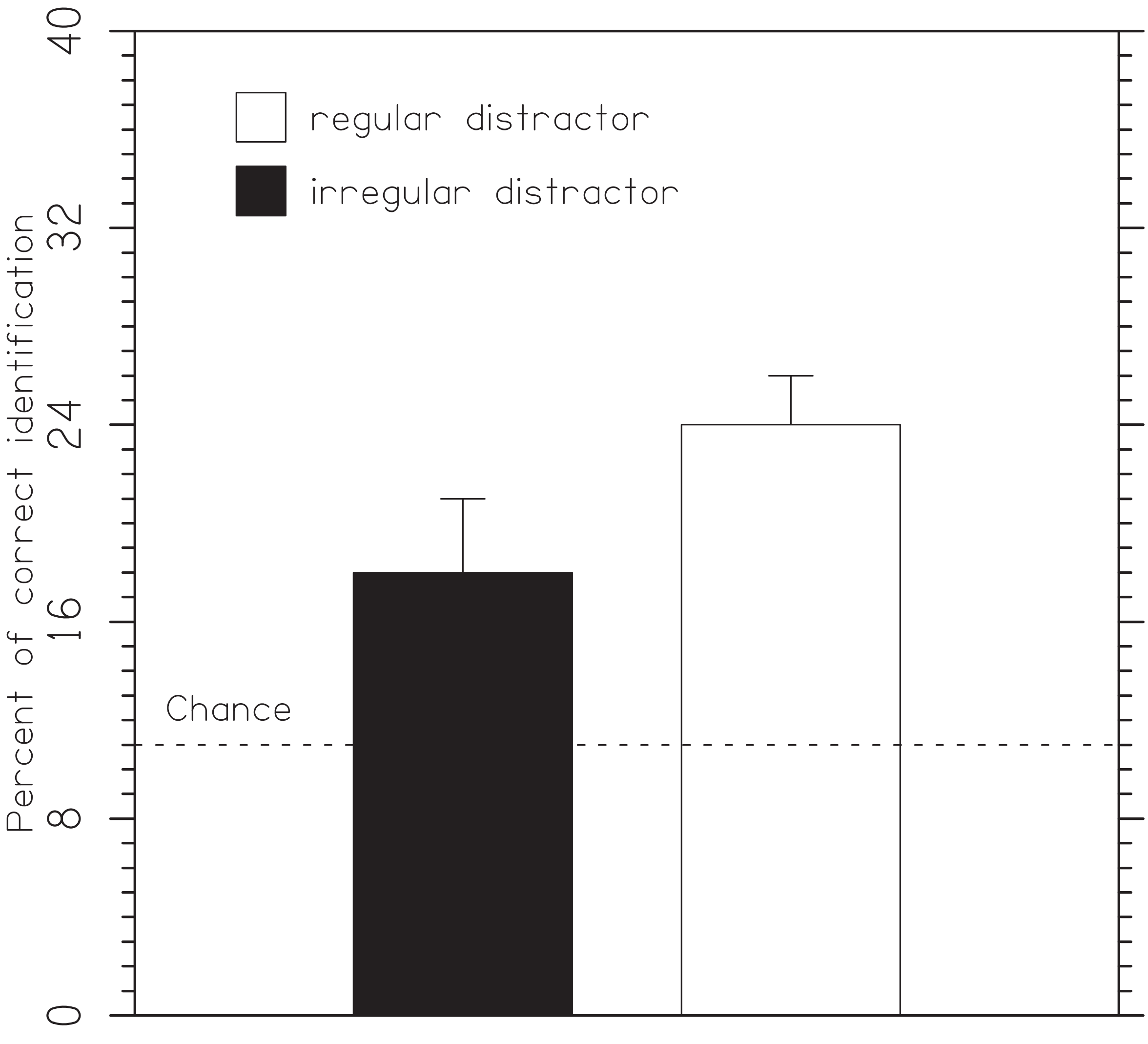
7. Au clair de la lune (m=73.8%, s.d=7.2%)

8. Ah vous dirais-je maman (m=70%, s.d=8.4%)




Irregular familiar and regular distractor melody
600
150
Pitch (log(Hz))
o
e
i
y
a
o
e
i
e
a
o
y
i
o
y
0
600
1200
1800
2400
time (ms)


Irregular familiar and distractor melody
600
150
Pitch (log(Hz))
o
e
i
y
a
o
e
i
e
a
o
y
i
o
y
0
600
1200
1800
2400
time (ms)



regular distractor
irregular distractor
Chance
Percent of correct identification
0
8
16
24
32
40