MELODIC FISSION IN TRANCE MUSIC

The Perception of Interleaved Vocal and Non-Vocal Melodies

By Lawrence Atherton

Abstract

While many studies have examined melodic fission of familiar and unfamiliar interleaved Western melodies, melodic fission in trance music, an electronic dance music genre, has not yet been studied. Melodic material is relatively constant throughout a trance song, while timbre, texture, and dynamics vary over time. In this study, participants listened to several clips of trance music with two or more competing melodic lines and evaluated which were melodic and harmonic. Answers were based primarily on how conjunct each line was, although some disjunct lines were segregated into two streams. Lyrical content, rhythmic simplicity, past musical training, familiarity with the genre, and connections drawn to other genres further affected the perception of melody in trance music.

I. Introduction

Listeners separate different sources from their aural environment into cohesive entities in a process called *stream segregation*, a term introduced by Bregman and Campbell (1971)[1]. These streams are organized mentally; they are not directly tied to physical properties of sound, although such properties may influence the streams via the percepts of timbre[i], pitch[ii], and rhythm[iii]. Dowling (1973)[2] called the process by which listeners segregate a pitch stream[iv] from a complex audio source[v] *melodic fission*, and the corresponding process for rhythms *rhythmic fission* (1968)[3]. Since rhythm and pitch are both integral components of melody, the combination of both processes can also be called *melodic fission*, although it is frequently abbreviated to *fission*.

Dowling (1973, Experiment II)[4] developed a system for testing the boundaries of fission in signals[vi] with two interleaved[vii] melodies wherein he played participants an unfamiliar melody (the *target*) then played it again, but interleaved with *distractor*[viii] tones. In some trials, the target melody changed between the first presentation and the interleaved presentation; participants were asked whether it had changed. In other experiments, Dowling presented a familiar target interleaved with a distractor and tested whether participants recognized the familiar melody. Most research involving fission from interleaved melodies uses methods like these, although some studies have deviated slightly; for example, Bey and McAdams (2003)[5] played the target after the interleaved signal in order to test post-recognition of unfamiliar melodies and remove the supposed benefit of hearing the target before the trial.

Pitch distance between two melodic streams is one of the major influences of fission. Miller and Heise (1950)[6] found that in a signal with two alternating pure tones[ix], listeners would segregate the signal into upper and lower streams when the frequency difference between the two tones was greater than fifteen percent, or three semitones[x]. To the contrary, Dowling (1973)[7] found that a mean frequency difference of six to twelve semitones between two melodies was necessary for fission to occur, depending on how conjunct[xi] or disjunct[xii] the melodies were. He attributed this difference to the fact that melodies may have many notes above or below their mean pitch. That is, two melodies with mean frequencies that are six semitones apart will invariably have many individual notes that are only three semitones apart and thus within the range found by Miller and Heise.

Bey and McAdams (2003)[8] expanded on these results to find that the optimum range of mean frequency distances for inducing fission is approximately ten to fifteen semitones, with best performance at twelve semitones. Bey and McAdams also found that participants performed worse in trials with distances of twenty-four semitones than in the zero-distance control trial; they attributed this result to the attentional distraction caused by the extreme pitch of one of the melodies. In an experiment using interleaved scales to test the effect of pitch distance on segregation, Gregory

---

1   A. S. Bregman and J. Campbell. "Primary auditory stream segregation and perception of order in rapid sequences of tones," Journal of Experimental Psychology 89 (1971): 244–249.

2   W. J. Dowling. "The perception of interleaved melodies," Cognitive Psychology 5 (1973): 322–337.

3   W. J. Dowling. "Rhythmic fission and perceptual organization," Journal of the Acoustical Society of America 44, no. 1 (1968): 369.

4   Dowling, "The perception of interleaved melodies."

5   C. Bey and S. McAdams. "Postrecognition of interleaved melodies as indirect measure of auditory stream formation," Journal of Experimental Psychology: Human Perception and Performance 29, no. 2 (2003): 267-279.

6   G. A. Miller and G. A. Heise. "The trill threshold," Journal of the Acoustical Society of America 22 (1950): 637–638.

7   Dowling, "The perception of interleaved melodies."

8   Bey and McAdams, "Postrecognition of interleaved melodies as indirect measure of auditory stream formation."

(1994)[9] found that participants frequently segregated the signal into upper and lower streams, rather than by scale, provided that there was no timbre difference between the scales. When he replaced the scales with disjunct melodies, nearly all participants segregated the signal this way.

Timbre is also a significant predictor of fission and may at times be more significant than pitch, provided that the timbre difference between two sources is large enough. Wessel (1979)[10] presented a repeated ascending line of three notes with timbre alternating between notes, and showed that when the timbral difference between adjacent notes was small, the repeated ascending lines were perceived, but when the timbral difference was large, the notes were grouped by timbre and two descending, repeated streams were perceived. This result is now called the Wessel Effect. To test the effect of timbre on segregation, Gregory (1994)[11] played interleaved scales and showed that with no timbre difference, participants segregated the signal into upper and lower streams, but as the timbre difference increased, segregation of full scales, that is segregation by timbre, became dominant. Bey and McAdams (2003)[12] replicated this result using their post-recognition experimental model, finding that the likelihood of fission increased with dissimilarity in timbre.

Rhythm can also influence a listener's ability to segregate two interleaved melodies. Studies of rhythm from a bottom-up[xiii] perspective have shown that participants' ability to integrate, or perceive as one stream, two single-frequency streams is dependent on pitch and not on regularity or irregularity of rhythm (van Noorden 1975; George and Bregman 1989)[13, 14]. Despite this finding, when Jones, Kidd, and Wetzel (1981)[15] studied rhythm from a top-down[xiv] perspective, they found that rhythm had a strong effect on fission from an acoustic mixture. Instead of using a target and distractor, they played two tones embedded in a distractor stream and asked participants to judge the order of the two tones. They found that the participants' performance on this task was weak when the tones had isochronous[xv] rhythms, and that performance increased as the tempo[xvi] difference between the tones increased. Devergie, Grimault, Tillmann, and Berthommier (2010)[16] experienced increased performance for fission of a rhythmically irregular target when using a rhythmically regular distractor instead of an irregular one. Therefore, a distractor may be more easily ignored if it is not isochronous with the target and has a regular rhythm.

However, a target will also be more easily segregated if its rhythm is not too complex. Essens (1986)[17] found that when participants tried to reproduce rhythms with non-integer subdivisions[xvii], they skewed the rhythms to subdivisions with integer relationships. Thus, while a distractor with simple rhythm aids fission, a target with a complex rhythm can be very difficult to segregate.

In a study of the musical features that aid melody identification, Schulkind, Posner, and Rubin (2003)[18] found that melody identification was strongest at phrase boundaries[xviii]. Their

---

9   A. H. Gregory. "Timbre and auditory streaming," Music Perception 12 (1994): 161–174.

10   D. L. Wessel. "Timbre space as a musical control structure," Computer Music Journal 3 (1979): 45–52.

11   Gregory, "Timbre and auditory streaming."

12   Bey and McAdams, "Postrecognition of interleaved melodies as indirect measure of auditory stream formation."

13   L. P. A. S. van Noorden. "Temporal coherence in the perception of tone sequences" (Ph.D. thesis, Eindhoven University of Technology, 1975).

14   M. F.-S. George and A. S. Bregman. "Role of predictability of sequence in auditory stream segregation," Perception & Psychophysics 46 (1989): 384-386.

15   M.R. Jones, G. Kidd, and Wetzel, R. 1981. Evidence for rhythmic attention. Journal of Experimental Psychology: Human Perception and Performance 7:1059–1073.

16   A. Devergie, N. Grimault, B. Tillmann, and F. Berthommier. "Effect of rhythmic attention on the segregation of interleaved melodies," Journal of the Acoustical Society of America 128, no. 1 (2010): EL1-EL7.

17   P. J. Essens. "Hierarchical organization of temporal patterns," Perception & Psychophysics 40 (1986): 67–68.

18   M. Schulkind, R. Posner, and D. Rubin. "Musical features that facilitate identification: How do you know it's "your" song when they finally play it?," Music Perception 21, no. 2 (2003): 217-249.

sources found that long notes and long rests create temporal accents (Handel 1989; Jones 1987; Jones and Boltz 1989)[19, 20, 21]. Schulkind et al. reasoned that while these types of durations do not encode any more information than other notes, listeners are more familiar with these notes because of the attention attracted by the resulting temporal accents. Also, Jones (1987)[22] and Narmour (1990)[23] found that most musical expectations are formed, met, and thwarted around phrase boundaries; Schulkind et al. attributed this finding to the high number of simultaneous musical accents[xix] around phrase boundaries, such as temporal accents and pitches close to the tonic. Schulkind et al.'s experiment validated these claims; melody identification for popular songs such as "Frosty the Snowman" was highest at phrase boundaries.

Familiarity with the material may give advantages to melodic fission. In a study on the effect of rhythm on segregation, Devergie et al. (2010)[24] found that participants' familiarity with the target melodies was the only consistent predictor of identification. They contrasted this with the result from Bey and McAdams' study (2002)[25], which reported that listeners segregate melodies stored in short-term memory by chance, given that there were no other acoustic cues for segregation (like pitch or timbre difference). Although listeners like melodies that they have heard more than once before more than they like previously unheard melodies (Peretz, Gaudreau, and Bonnel 1998; Halpern and Müllensiefen 2008; Weiss 2011)[26, 27, 28], there are no significant correlations between how much a new melody is liked and how well it is remembered (Weiss 2011)[29]. However, liking a piece of music may lead to repeated listening. This repetition would move the music into long-term memory and then confer an advantage to fission, as in Devergie et al. (2010)[30]. Also, a listener who has a strong familiarity with a body or genre of music often perceives it differently (Dowling 1973)[31].

Another important quality of a musical signal is its vividness, or how much attention it attracts and how memorable it is. Dowling (1973)[32] points out that listeners pay attention to vivid stimuli, and while loudness, brightness[xx], and other percepts can contribute to vividness, associations with knowledge of existing genres of music (whether of the same genre as the stimulus or not) may also increase vividness. Listeners also have a higher mental response to vocal timbres, which are familiar because of how much the voice is used in everyday life. An fMRI study by Belin, Zatorre, and Ahad (2002)[33] showed more brain activity in the presence

19   S. Handel. Listening: An introduction to the perception of auditory events (Cambridge, MA: MIT Press, 1989).

20   M.R. Jones. "Dynamic pattern structure in music: Recent theory and research," Perception & Psychophysics 41 (1987): 621–634.

21   M.R. Jones and M. Boltz. "Dynamic attending and reactions to time," Psychological Review 96 (1989): 459–491.

22   Jones, "Dynamic pattern structure in music: Recent theory and research."

23   E. Narmour. The analysis and cognition of basic melodic structures: The implication-realization model (Chicago: University of Chicago Press, 1990).

24   Devergie et al., "Effect of rhythmic attention on the segregation of interleaved melodies."

25   C. Bey and S. McAdams. "Schema-based processing in auditory scene analysis," Perception & Psychophysics 64 (2002): 844–854.

26   I. Peretz, D. Gaudreau, and A.-M. Bonnel. "Exposure effects on music preference and recognition," Memory and Cognition 26 (1998): 884-902.

27   A. R. Halpern and D. Müllensiefen. "Effects of timbre and tempo change on memory for music," The Quarterly Journal of Experimental Psychology 61 (2008): 1371-1384.

28   M. W. Weiss. "Vocal timbre influences memory for melodies" (Master's thesis, University of Toronto, 2011).

29   Ibid.

30   Devergie et al., "Effect of rhythmic attention on the segregation of interleaved melodies."

31   Dowling, "The perception of interleaved melodies."

32   Ibid.

33   P. Belin, R. J. Zatorre, and P. Ahad. "Human temporal-lobe response to vocal sounds," Cognitive Brain Research 13, no. 1 (2002): 17-26.

of vocal audio than non-vocal audio and more brain activity in the presence of vocal signals with words than those non-verbal expressions, like laughter. Weiss (2011)[34] confirmed that vocal melodies are remembered better than other familiar timbres and unfamiliar timbres. Finally, listeners have a better memory for melodies played on an instrument with which they have trained (Tervaniemi, Rytkönen, Schröger, Ilmoniemi, and Näätänen, 2001)[35].

II. A Brief Aside on Form in Trance Music

The form of trance music is much more rigidly structured than other genres. Popular trance music from at least as far back as 2000 almost always has an intro, a breakdown, a buildup, a takeoff/anthem, and an outro. These sections guide the tension of the song, which climbs slowly through the intro, drops off at the breakdown, and climbs during the buildup back to its peak at the takeoff before winding down through the outro. The most salient melody of a trance song is usually not introduced in its full form until the breakdown, buildup, or takeoff; this structure means that the first few minutes of the song contain percussive and harmonic content but little melodic content.

Trance melodies are somewhat unique because the properties required for melodic identification—intervals and rhythms—do not change between phrases. Trance music induces tension through the form described above by changing dynamics, rhythmic activity, timbre, and textural density over time (Iler 2011).[36] The phrase boundaries of trance music do not have significant melodic elements like cadences, but important variations in timbre, dynamics, and texture co-occur at phrase boundaries. Thus, these melodies differ greatly from the Western ones used in other studies and make trance songs a worthwhile body of music to study.

III. Methods

A. Participants

Three participants were selected for this study. All three participants (herein P1, P2, and P3) had more than ten years of musical training. P1 was trained on the trumpet in ensemble and solo settings and on the voice in ensemble and scat singing. P2 was trained on the voice in ensemble and solo settings, but she had no training in scat singing or any other syllabic solo system. P3 was trained on the piano in a solo setting. Participant 1 was intimately familiar with the genre and the songs presented, while participants 2 and 3 were unfamiliar with the genre and with all songs presented.

---

34   Weiss, "Vocal timbre influences memory for melodies."

35   M. Tervaniemi, M. Rytkönen, E. Schröger, R. J. Ilmoniemi, and R. Näätänen. "Superior formation of cortical memory traces for melodic patterns in musicians," Learning and Memory 8 (2001): 295-300.

36   D. Iler. "Formal devices of trance and house music: Breakdowns, buildups, and anthems" (Master's thesis, University of North Texas, 2011).

B. Stimuli

Clips from ten trance songs were used. Two clips were taken from each song. The first clip was a sixteen-bar phrase from the buildup or the takeoff that contained multiple melodic lines in the same register with a similar loudness. Each of these phrases usually contained one or two full repetitions of all lines involved. The second clip, called the "context" clip, consisted of the previous clip and one to three phrases before it. In general, the guideline for choosing the start of the context clip was to extend backward until only one line was playing, but in some cases this was impossible. For example, in "Gravity," both melodic lines under question enter at the same phrase boundary, so the context clip simply extends one phrase before both lines enter. The reason for including the context clips was to observe how phrase boundaries informed the participants, and whether information from phrase boundaries would alter perceptions of melody. The trance songs were chosen from one music label (Anjunabeats) from the years 2004-2007 for the sake of consistency and because many popular songs from these years fulfill the requirements of multiple simultaneous competing melodic lines.

C. Procedure

Participants were tested alone. The entire test lasted about one hour. Before the start of the test, the participant was told that they should listen for the melody, or the part they would sing or dance to, in the music they were about to hear. They were told that the experiment had no wrong answers and were encouraged to share any thoughts they had, especially if they were unsure. For each trial, the participant first heard the one-phrase clip on repeat, and they were encouraged to listen for as long as they needed before they made a decision in order to minimize the impact of stress and short-term memory on their answer. Next, the music was stopped, and the participant communicated the melodic and harmonic lines they had chosen by describing the timbres of the lines or by trying to sing the notes of the lines. If clarification was needed, the clip was started again so that the participant could point out which lines they had chosen. All thoughts expressed aloud by the participant were then recorded in writing.

　　　Next, the participant was told that they would hear the same clip with some context before it, and they were encouraged to describe any thoughts about the melody without waiting until the end of the clip. The context was played once through. Any thoughts the participant expressed aloud during or after the context was played were recorded. Participants were not told any information about the names of the songs or the artists who produced them until the end of the test.

IV. Results

In Table 1, the code name for each song is used. For full song titles, see Appendix 1. Appendices 2-5 contain graphs of the interval and tatum[xxi] distributions of several different groups of lines: non-vocal and vocal, melodic and harmonic, etc. In Table 1, each cell contains the source(s) that the participant identified as melodic followed by the source(s) identified as harmonic[37]. For

---

37   Sound files are hosted on the website soundcloud.com; the left-hand column of Table 1 contains the links to all sound files.

"Believe," all three lines are synths, so "Upper, Lower, Pad"[xxii] are used. For "Probspot," all three lines have the same timbre, so "Upper, Middle, Lower" are used. If only the upper notes of a disjunct line are perceived, it is written in italics.

TABLE 1
Melodies and harmonies
chosen by the participants

	Participant 1	Participant 2	Participant 3
"Amsterdam"	Voice, Synth	Voice[a], Synth	Synth, Voice[b]
Context	Voice, Synth	Synth, Voice	Synth, Voice
"Believe"	Upper, Lower	Pad, Upper	None, Upper
Context	Upper, Lower	Pad, Upper	*Lower*, Upper
"Dawn"	Piano, Synth	Piano, Synth	Synth, None[c]
Context	Piano, Synth	Piano, Synth	Synth, None
"Gravity"	Voice, Piano	Piano, Voice	Piano, Voice
Context	Voice, Piano	Piano, Voice	Piano, Voice
"Helsinki"	Voice, Synth	Synth + Voice, None	None[d], Voice
Context	Voice, Synth	Synth, Voice	Synth, Voice
"No One"	Voice, Synth	Voice, Synth	Voice, Synth
Context	Voice, Synth	Voice, Synth	Voice, Synth
"Probspot"	Middle, Lower	None, Upper + Middle	Upper + Middle, None
Context	Middle, Lower	Lower[e], Upper[f]+ Middle	Upper + Middle, None
"Surrender"	Voice, Synth	Synth+ Voice[f], None	*Synth*, Voice
Context	Voice, Synth	Synth, Voice	*Synth*, Voice
"Suru"	Voice, Synth	Voice[g], None	*Synth*, Voice
Context	Voice, Synth	Voice, Synth	*Synth*, Voice
"Won't Sleep"	Guitar, Synth	Guitar, Voice	Synth, Voice
Context	Guitar, Synth	Synth[h], Voice	*Guitar*, Voice

a. Liked the voice as a candidate but later decided it was too "ethereal."

b. Also identified the voice as too ethereal.

c. Could not hear the piano even if it was sung to her.

d. Could not hear the synth when the voice was present, didn't like the voice for melody.

e. Associated Lower with classical form ("it's like a Bach Toccata!").

f. Liked voice as a candidate but felt it was too "ethereal."

g. Did not know it was a voice.

h. With guitar "accenting notes from the synth."

V. Discussion

A. Fission in Non-Vocal Lines

Participants generally split lines with non-vocal timbres into two groups: melodic lines consisting mostly of unisons and second intervals (Appendix 2a) and harmonic lines with more perfect fifth, perfect fourth, and third intervals (Appendix 2b). This division corresponds roughly to conjunct (melodic) and disjunct (harmonic) lines. However, several factors overwhelmed this division, including timbre (P1: "Won't Sleep"), vividness related to both connections drawn to familiar music and to attentional distractors (P2: "Probspot", P3: "Dawn"), and stream segregation based on pitch distance (P3: "Believe"). Deviations like these diluted the lines' interval frequencies away from starkly large frequencies of major second intervals for Appendix 2a and perfect fifth intervals for Appendix 2b. In these appendices, the respective intervals are more frequent than the rest of the intervals, but not overwhelmingly so.

B. "Believe": Segregations of Disjunct Harmonic Lines Create Unison-Dominant Melodies

In the "Believe," "Surrender," "Suru," and "Won't Sleep" trials, P3 did not perceive lines strictly by timbre. Instead, she perceived only the upper notes of a disjunct harmonic line. The harmonies that P3 segregated had a higher percentage of perfect fifth intervals (30 percent, Appendix 3b) than the average harmonic line (20 percent, Appendix 2b), which is the likely reason they were segregated into upper and (non-perceived) lower streams. Also, the upper-stream melodies were much more conjunct (89 percent unison and second intervals, Appendix 3a) than the average perceived melody (54 percent unison and second intervals, Appendix 2a). Given these statistics and given that listeners generally segregate disjunct melodies into upper and lower streams when timbre differences are small (Gregory 1994)[38], it is strange that P1 and P2 segregated by timbre instead of interval size, perceiving less conjunct melodies than P3.

The answer to this discrepancy may lie in the three participants' musical training. P1 and P2 were trained in ensemble settings, where musicians had to keep track of how their own line fit into other musicians' parts; in contrast, P3 was trained in a solo piano setting, where there is no difference in timbre from note to note and where melody is often constructed from the highest notes of all the notes being played. This hypothesis is further reinforced by P2's response for the "Believe" trial. P2 chose a background line (the "pad" or soft chords) instead of one of the two more salient lines for the melody, and this pad line played exactly the same notes as P3's segregated line, albeit with less rhythmic variety. Thus, the urge to find a very conjunct melody was present, but P2's ensemble training kept her from segregating lines into upper and lower streams in order to create one.

C. "Probspot": Vividness and Drawing Connections to Familiar Music

In the "Probspot" trial, P2 was unsatisfied with the upper and middle lines in the first playing of the clip and could not hear the lower line. However, upon hearing the lower line playing by itself in the context, she immediately exclaimed, "That. That's the melody. It's like a Bach Toccata!"

---

38   Gregory, "Timbre and auditory streaming."

Tenuous as this association may be, it allowed her to keep track of the lower line even as the music continued through the original, dense clip section, where she had not heard the lower line on her first listen. P2's reaction to this line was the strongest of any of the trials, and also the only instance of the context clip allowing a participant to hear another line, so it is possible that associating unfamiliar music with familiar music facilitates fission. Bey and McAdams (2002)[39] found that short-term memory gave no aid to fission, but it seems from this study that a melody stored in short-term memory can be more easily segregated when it is linked to a different long-term memory.

D. "Dawn": Vividness and Attentional Distractors

In the "Dawn" trial, P3 was unable to hear the piano line, even if the examiner sang along to it while the clip was playing. The synth line of "Dawn" is at times an octave above the piano line. Also, the piano line is only present when a percussion track, a bass line, and a mid-bass line are also competing for attention. Thus, it is possible that these attentional distractors, such as pitch height, loudness, and rhythmic variation, detracted so much from the vividness of the piano line that P3 was unable to hear it. While Bey and McAdams (2003)[40] found that a mean frequency difference of two octaves was enough to impede fission from two interleaved sources, this study shows that attentional distraction to the point of inaudibility is possible even with a mean frequency difference of less than an octave, provided that other material is present.

E. "Won't Sleep": Familiarity with the Material Affects Timbre Preference

In the "Won't Sleep" trial, P1 chose the disjunct guitar melody over the conjunct synth melody, while P2 chose the synth melody and P3 chose a conjunct segregation of the guitar melody. The synth melody has a timbre (sawtooth) and texture (oscillating or "ducking" loudness) that is common in trance music; for example, the "Helsinki" trial also uses this synth. Since timbre dissimilarity increases fission ability (Gregory 1994; Bey and McAdams 2003)[41, 42], it is possible P1 compared the synth line in "Won't Sleep" to others stored in his long-term memory, then found the guitar more dissimilar and interpreted the guitar as the melody. P2 and P3, then, would still choose conjunct lines, picking the synth or the upper notes of the guitar, which contain mostly second intervals.

F. Fission in Vocal Lines

The interval distribution for vocal melodies chosen by the participants resembles that of the non-vocal melodies: the distribution has a high frequency of major seconds, and 78 percent of intervals are conjunct (Appendix 4a). However, interval distribution for the vocal harmonies does not differ significantly from that of the vocal melodies: major seconds are still most common, and 63 percent of intervals are conjunct (Appendix 4b). This lack of difference is probably due to the fact that there are no vocal lines with many disjunct intervals in any of the songs used.

---

39   Bey and McAdams, "Schema-based processing in auditory scene analysis."
40   Bey and McAdams, "Postrecognition of interleaved melodies as indirect measure of auditory stream formation."
41   Gregory, "Timbre and auditory streaming."
42   Bey and McAdams, "Postrecognition of interleaved melodies as indirect measure of auditory stream formation."

Conjunct vocal melodies are the norm in must current genres of music, and the vocal lines were consistently conjunct. Why then did P2 and P3 identify vocal melodies in so few of the trials? The answer has partly to do with lyrical content and partly to do with whether the lyrical lines are simple and have repetitive rhythms.

G. *"Gravity," "Helsinki," "Surrender," and "Suru": Lack of Lyrical Content Influences Fission*

The only trial in which all three participants agreed upon one melody was "No One," which has lyrics. P1 also responded that all non-lyric vocal lines were melodies ("Gravity," "Helsinki," "Surrender," and "Suru"). Conversely, P3 rejected all non-lyric vocal lines, and P2 only chose one when she thought its timbre was non-vocal ("Suru"). P2 and P3's harmonic interpretations of non-lyric vocal lines reinforce Belin et al. (2002)[43]'s results, which showed that vocal sounds with words elicited a greater brain response than those without words.

However, Belin et al.'s study does not explain why P1 would have such a strong response to non-lyric vocal melodies. P1's training on the "instrument" of scat singing may have afforded him a better memory for melodies sung syllabically, as per the results of Tervaniemi et al. (2001)[44]. It is also possible that P1's extreme familiarity with the songs gave him the long-term memory advantage described by Devergie et al. (2010)[45]. It is impossible to know which, if either, affected P1's responses without isolating scat singing and genre familiarity. Nonetheless, it is not probable for a listener who is both unfamiliar with trance and untrained in syllabic solo singing to segregate a non-lyric trance melody.

H. *"No One" and "Amsterdam": Rhythmic Simplicity and Repetition Improve Fission*

The participants only unanimously chose the vocal line as the melody in one of the three trials with lyrics ("No One"). No participant chose the vocal line of "Won't Sleep," probably because this trial has two lines in a higher register also competing for attention and acting as attentional distractors. However, the other lyric trials, "No One" and "Amsterdam," feature female voices in the same register as repetitive disjunct synth lines, and repetitive disjunct lines can be easily ignored when acting as distractors (Devergie et al. 2010)[46]. Why, then, did P2 and P3 describe the voice in "Amsterdam" as too "ethereal" but accept the voice in "No One"? To be fair, there are several measures in the "Amsterdam" trial where the voice simply isn't present. In other measures, the voice is reversed, making it effectively a syllabic line for those measures.

However, the second half of the "Amsterdam" clip contains a lyric, non-reversed voice. Still, there are a few key differences between this voice and the voice heard in "No One." The rhythm and lyrics of "No One" repeat halfway through the phrase, and listeners like melodies they have heard at least once before more than completely novel ones (Weiss 2011)[47]; it is possible that a high preference for the voice of "No One," resulting from its repetition, led P2 and P3 to report it over the synth line. "Amsterdam," on the other hand, does not have rhythms or lyrics that repeat within the clip.

"No One" also has a much simpler rhythm than "Amsterdam." The rhythm of the vocal line

43   Belin et al., "Human temporal-lobe response to vocal sounds."
44   Tervaniemi et al., "Superior formation of cortical memory traces for melodic patterns in musicians."
45   Devergie et al., "Effect of rhythmic attention on the segregation of interleaved melodies."
46   Ibid.
47   Weiss, "Vocal timbre influences memory for melodies."

in "No One" places emphasis on downbeats 1, 3, and 4 of the measure (Appendix 5a), whereas the rhythm of the vocal line in "Amsterdam" does not emphasize any eighth note much more than the others (Appendix 5b). The tatum distribution[xxiii] of the vocal line in "No One" leads clearly to the start of each measure, whereas the tatum distribution of the vocal line in "Amsterdam" is disorienting because it does not seem to lead anywhere.

Finally, in the clip for "No One," the vocal note onsets are short and clearly defined, but in "Amsterdam," the vocal note onsets are long and unclear, making the rhythm seem even more irregular. Since non-integer rhythm ratios are not easily reproduced (Essens 1986)[48], the perceived rhythmic distortion caused by long note onsets could make the vocal line in "Amsterdam" difficult to follow. All of these rhythmic complications could have caused P2 and P3 to report the vocal line of "No One" but not that of "Amsterdam."

Despite all this, P1 still reported the vocal line of "Amsterdam" as the melody. It was likely that he had the rhythm encoded very well in his long-term memory, given that he sang along without any errors and reported a strong preference for the song. This familiarity would nullify any of the negative effects of irregular rhythm on melodic perception that P2 and P3 experienced.

I. Phrase Boundaries and Familiarity with the Genre

While participants' answers rarely changed when shown the context clip, the participants did make some remarks about what they expected to hear from phrase boundaries. P1 expected new material at a phrase boundary to act melodically, while P2 and P3 remarked in some of the trials that new material reinforced old material. Indeed, P1 only identified one line that played by itself in the context phrase as a melody ("Amsterdam"), whereas P2 and P3 picked such phrases at chance levels (5/10 trials each). P1's expectations follow naturally from his familiarity with the genre; in trance music, the introduction of the most salient melody is often prolonged until the tension increases in the buildup and takeoff. P2 and P3, who were unfamiliar with the genre, gleaned no information from phrases and phrase boundaries.

VI. Conclusions

While a variety of factors influenced melodic fission in the complex acoustic signal of a clip from a trance song, participants generally identified more conjunct lines as melodies and more disjunct lines as harmonies. Vocal lines with lyrics and simple rhythms were universally considered melodies, whereas participants unfamiliar with the genre had trouble segregating lyric lines with irregular rhythms and identified non-lyric lines as harmonic. Occasionally, solo-trained participants segregated very disjunct lines into upper and lower streams, ignoring timbre cues. Familiarity with the genre affected timbre preference and significantly increased the number of decisions that correlated with phrase boundaries. Vividness from association with any familiar material aided the ability to segregate melodic lines, and decreased vividness from attentional distractions hindered it.

Future studies should attempt to isolate the influences of genre familiarity, song familiarity, and training in syllabic (i.e. non-lyric) singing, as well as test the correlation between solo polyphonic training and upper/lower segregation of disjunct lines. Testing more lyric material

---

48   Essens, "Hierarchical organization of temporal patterns."

will reveal which rhythmic factors (regularity, repetition, note onset length, etc.) influence fission of vocal lines the most. The elusive problem of vividness, or improved segregation ability from drawing mental connections between unfamiliar and familiar material, could be studied by testing participants' segregation ability in trance remixes of songs they are familiar with. Finally, future research should examine whether the association with conjunct lines and melody (and correspondingly, disjunct lines and harmony) extends to less melodic genres of electronic dance music.

Bibliography

Belin, P., Zatorre, R. J., and Ahad, P. 2002. Human temporal-lobe response to vocal sounds. Cognitive Brain Research 13, no. 1:17-26.

Bey, C., and McAdams, S. 2002. Schema-based processing in auditory scene analysis. Perception & Psychophysics 64:844–854.

Bey, C., and McAdams, S. 2003. Postrecognition of interleaved melodies as indirect measure of auditory stream formation. Journal of Experimental Psychology: Human Perception and Performance 29, no. 2:267-279.

Bregman, A. S., and Campbell, J. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology 89:244–249.

Devergie, A., Grimault, N., Tillmann, B., and Berthommier, F. 2010. Effect of rhythmic attention on the segregation of interleaved melodies. Journal of the Acoustical Society of America 128, no. 1:EL1-EL7.

Dowling, W. J. 1968. Rhythmic fission and perceptual organization. Journal of the Acoustical Society of America 44, no. 1:369.

Dowling, W. J. 1973. The perception of interleaved melodies. Cognitive Psychology 5:322–337.

Essens, P. J. 1986. Hierarchical organization of temporal patterns. Perception & Psychophysics 40:67–68.

George, M. F.-S., and Bregman, A. S. 1989. Role of predictability of sequence in auditory stream segregation. Perception & Psychophysics 46:384-386.

Gregory, A. H. 1994. Timbre and auditory streaming. Music Perception 12:161–174.

Halpern, A. R., and Müllensiefen, D. 2008. Effects of timbre and tempo change on memory for music. The Quarterly Journal of Experimental Psychology 61:1371-1384.

Handel, S. 1989. Listening: An introduction to the perception of auditory events. Cambridge, MA: MIT Press.

Huron, D. "Music Cognition Handbook: A Dictionary of Concepts." Last modified 2000. http://csml.som.ohio-state.edu/Resources/Handbook/

Iler, D. 2011. "Formal devices of trance and house music: Breakdowns, buildups, and anthems." Master's thesis, University of North Texas.

Jones, M. R. 1987. Dynamic pattern structure in music: Recent theory and research. Perception

& Psychophysics 41:621–634.

Jones, M. R., and Boltz, M. 1989. Dynamic attending and reactions to time. Psychological Review 96:459–491.

Jones, M. R., Kidd, G., and Wetzel, R. 1981. Evidence for rhythmic attention. Journal of Experimental Psychology: Human Perception and Performance 7:1059–1073.

Miller, G. A., and Heise, G. A. 1950. The trill threshold. Journal of the Acoustical Society of America 22:637–638.

Narmour, E. 1990. The analysis and cognition of basic melodic structures: The implication-realization model. Chicago: University of Chicago Press.

Peretz, I., Gaudreau, D., and Bonnel, A.-M. 1998. Exposure effects on music preference and recognition. Memory and Cognition 26:884-902.

Schulkind, M., Posner, R., and Rubin, D. 2003. Musical features that facilitate identification: How do you know it's "your" song when they finally play it?. Music Perception 21, no. 2:217-249.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., and Näätänen, R. 2001. Superior formation of cortical memory traces for melodic patterns in musicians. Learning and Memory 8:295-300.

van Noorden, L. P. A. S. 1975. "Temporal coherence in the perception of tone sequences." Ph.D. thesis, Eindhoven University of Technology.

Weiss, M. W. 2011. "Vocal timbre influences memory for melodies." Master's thesis, University of Toronto.

Wessel, D. L. 1979. Timbre space as a musical control structure. Computer Music Journal 3:45–52.

Endnotes

i   **Timbre**, "tone color," or "tone quality" are catch-all terms that denote those properties of a sound – other than pitch and loudness – which combine to produce an overall auditory identity or character. The notion of timbre is closely associated with the "identifiability" or "distinguishability" of a sound, or class of sounds. Musicians will thus speak of the timbre of a violin, or the class of "brassy" timbres. ("Music Cognition Handbook: A Dictionary of Concepts," David Huron, last modified 2000, http://csml.som.ohio-state.edu/Resources/Handbook/)

ii   **Pitch** is a psychological/musical term denoting the mental correlate of frequency ("Music Cognition Handbook").

iii   A **rhythm** is a pattern of durations that is usually characterized by relatively strong and weak beats.

iv   A stream is the auditory experience of a "line of sound" ("Music Cognition Handbook"). A **pitch stream** is thus a sequence of pitches or notes played in succession without any other competing audio. In experiments, a pitch stream often does not have enough variance in timbre and/or rhythm to be considered a melody.

v   Examples of **complex audio source**s: a song recorded in a studio, or two people singing over each other; any combination of multiple competing or parallel streams that do not totally blend into one another.

vi   A **signal** is any individual stream or complex audio source.

vii   Two pitch streams are **interleaved** when they are played simultaneously, with one pitch from one stream followed by one pitch from the other.

viii   **Distractor tones** are a pitch stream designed to distract the listener from a melody. If played alone, distractor tones usually do not seem melodic in the traditional sense.

ix   A **pure tone** is made of only one frequency; i.e., a sine wave. Contrasted with a complex tone, which is comprised of several pure tones. All instruments create complex tones, but a pure tone arises from a tuning fork and from whistling.

x   A **semitone**, also called a "half-step," represents the smallest pitch distance between notes in Western music.

xi   **Conjunct** intervals are also called "steps." These intervals correspond to a distance of no semitones (unison), one semitone (minor second), or two semitones (major second) between adjacent notes. In some circles, unison intervals are not called steps, but the phrase conjunct interval refers to both unison intervals and steps.

xii   **Disjunct** intervals are also called "skips" or "jumps." Any interval larger than a conjunct interval is disjunct.

xiii   Studying a phenomenon with a "**bottom-up**" perspective means analyzing the physical attributes of sound: amplitude, frequency, time, to name a few.

xiv   **Top-down** analysis focuses on psychological percepts: loudness, pitch, and perceived duration, to name a few.

xv   **Isochronous** rhythms are well-synced and have similar tempos.

xvi   **Tempo** is a measure of the speed of a stream. It is usually measured in beats per minute.

xvii   The tactus, or perceived basic pulse or beat ("Music Cognition Handbook"), is usually divided into integer subdivisions; for example, it can be divided in three to create triplets, or four to create sixteenth notes. A **non-integer subdivision** cannot be divided into whole number divisions and would be more difficult to perceive, such as a $1/2.5^{th}$ note.

xviii   Music is generally organized by grouping notes into phrases, which often contain a complete melodic idea, or half of a melodic idea. The **phrase boundary** is the boundary between such phrases.

xix   A **musical accent** emphasizes a note in some manner. This may be by loudness (a "dynamic accent," widely recognized as a > in scores), in increased duration (an "agogic accent"), or by virtue of being high-pitched (a "tonic accent"). Some notes may be accented without having any of these qualities; they may be close to the *tonic* note (the harmonic function associated with the most stable scale degree ("Music Cognition Handbook")), or they may fall on a strong beat of the music's *meter* (an organization into strong and weak beats; unlike rhythm, meter is usually simple and applies to a whole song).

xx   A complex tone is **brighter** if it has more, higher, sinusoidal components, or overtones. An example of an instrument that is bright (has many overtones) when played at a loud volume is a trumpet. Conversely, a flute has few overtones and is not as bright. Brightness is a relative term, although it may be measured via the spectral centroid, which is an absolute measure.

xxi   A **tatum** is a "temporal atom," the shortest duration in a notated musical work that can be used as a divisor for all other durations. For example, if all nominal durations in a work are divisible into sixteenth durations, and the sixteenth duration is the largest such divisor, the sixteenth value is deemed the tatum for the work. The term tatum was coined at the Center for New Music and Audio Technologies at the University of California, Berkeley in 2000, and was named to evoke the rapid-fire piano playing of jazz keyboardist, Art Tatum ("Music Cognition Handbook").

xxii   A **pad** is a chord line in the background.

xxiii   A **tatum distribution** maps out how frequently a note occurs on each tatum in an individual measure. It shows which tatums are, on average, more emphasized than others in a given piece of music.

Appendices

Data for Appendices 2-5 was gathered by transcribing the trial songs' lines into midi files and running these files through a custom melody analysis program written in C++.
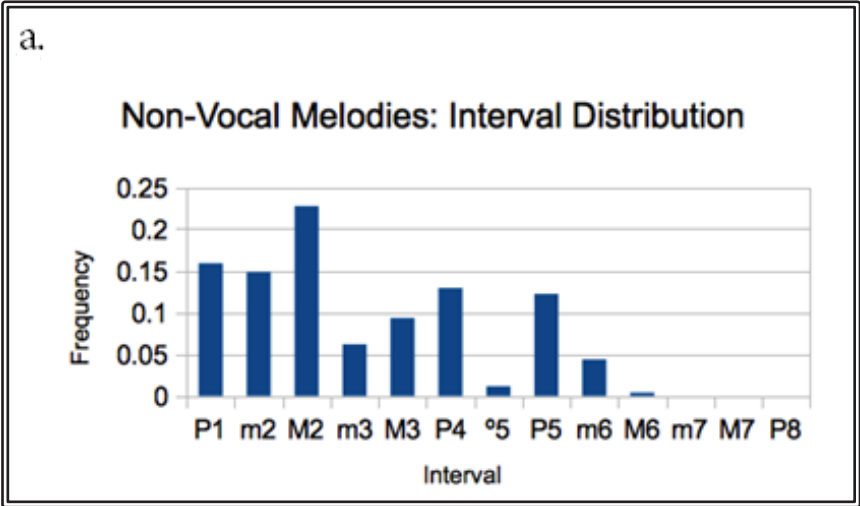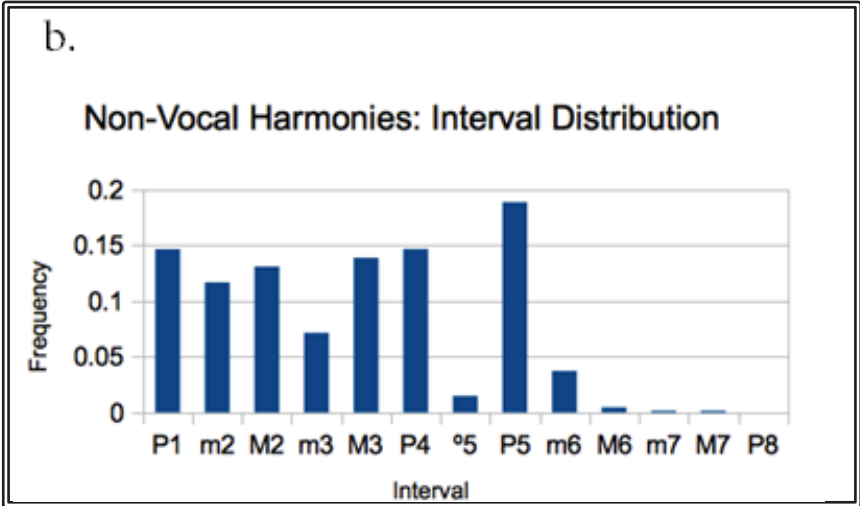
APPENDIX 1
LIST OF SONGS USED

Code Name	Artist – Song (Remixer if Applicable)
Amsterdam	Luminary – "Amsterdam" (Smith & Pledger Remix)
Believe	Smith & Pledger – "Believe"
Dawn	Super8 – "Dawn"
Gravity	P.O.S. – "Gravity"
Helsinki	Super8 & Tab – "Helsinki Scorchin'"
No One	Above & Beyond feat. Zoë Johnston – "No One on Earth"
Probspot	Endre – "I Kill for You" (Probspot Remix)
Surrender	Tranquility Base – "Surrender"
Suru	Super8 & Tab – "Suru"
Won't Sleep	Super8 & Tab – "Won't Sleep Tonight"

APPENDIX 2
INTERVAL DISTRIBUTIONS
OF NON-VOCAL SONGS

a.

Non-Vocal Melodies: Interval Distribution

b.

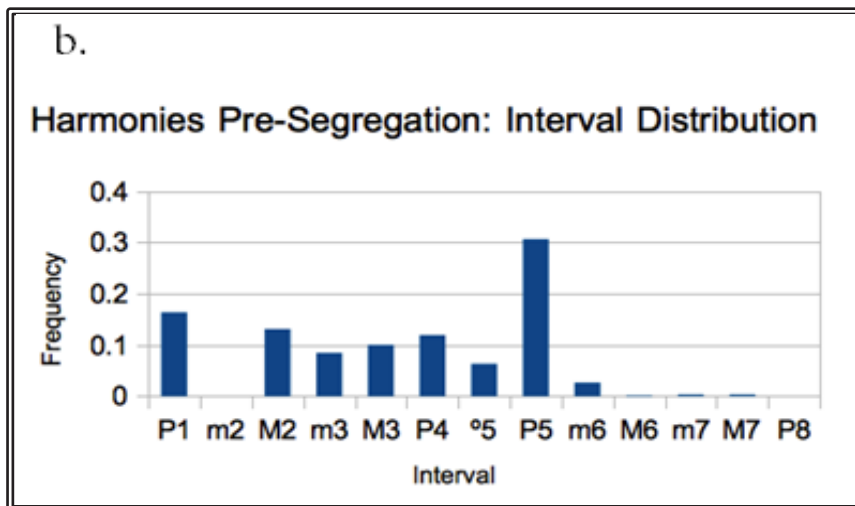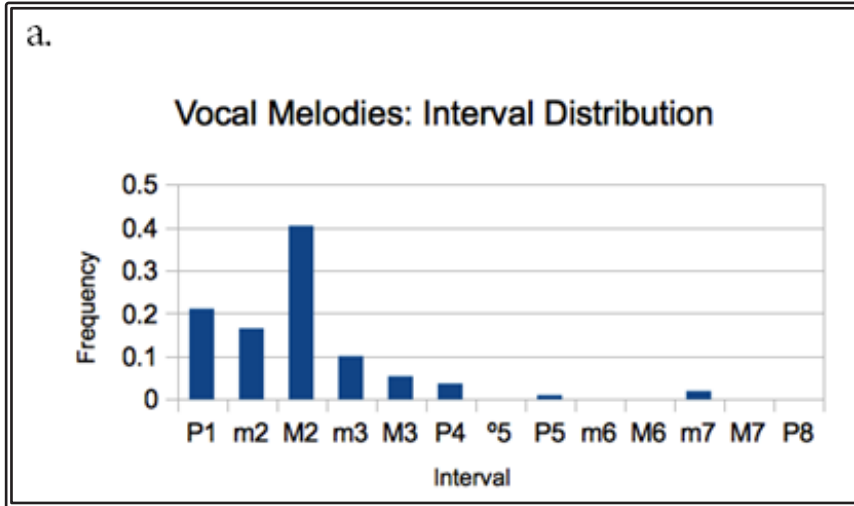**Non-Vocal Harmonies: Interval Distribution**

APPENDIX 3
Interval Distributions of Harmonies
Segregated into Melodies

a.

Segregated Melodies: Interval Distribution

b.

Harmonies Pre-Segregation: Interval Distribution
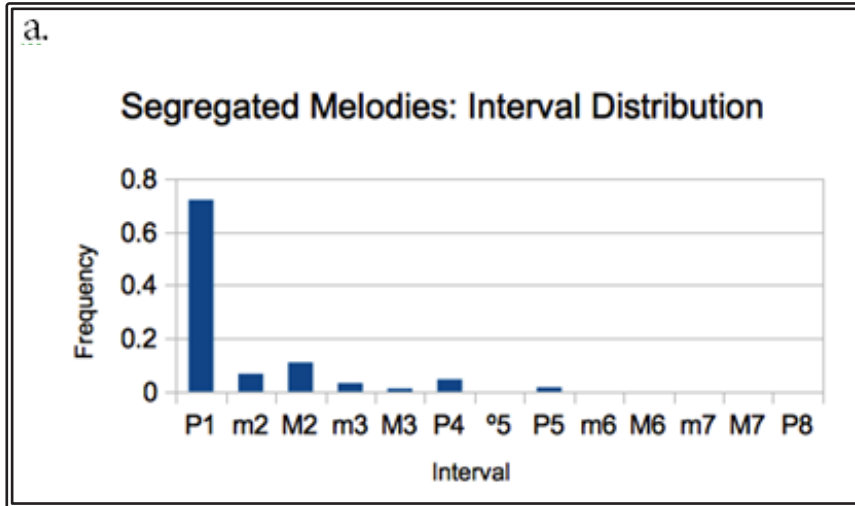
APPENDIX 4
Interval Distributions
of Vocal Songs

a.

Vocal Melodies: Interval Distribution

b.

Vocal Harmonies: Interval Distribution
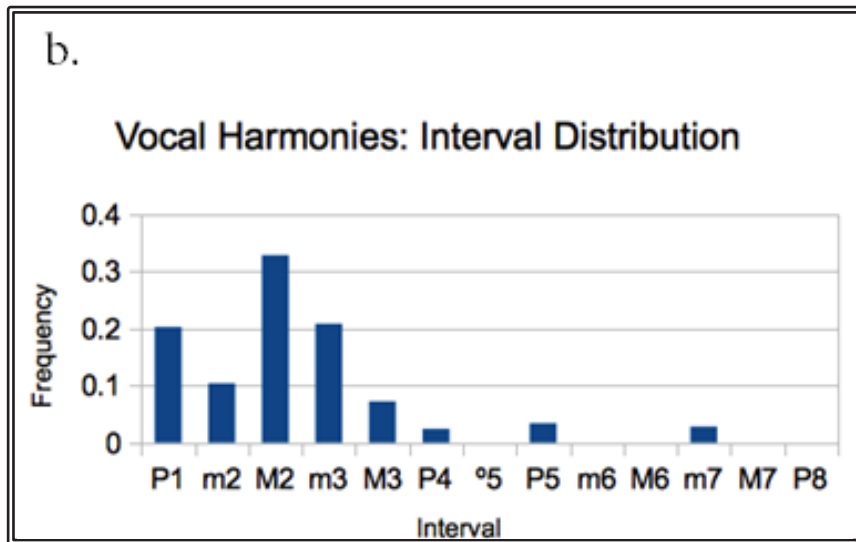
APPENDIX 5
Tatum Distributions of Lyric Lines

a.

Lyric Melody (No One On Earth): Tatum Distribution

b.

Lyric Harmony (Amsterdam): Tatum Distribution