

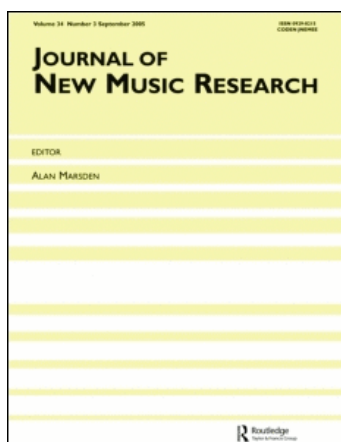
This article was downloaded by: [Ingenta Content Distribution - Routledge]

On: 10 March 2009

Access details: Access Details: [subscription number 791963552]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of New Music Research

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713817838>

Scanning the Dial: The Rapid Recognition of Music Genres

Robert O. Gjerdingen ^a; David Perrott ^b

^a Northwestern University, USA ^b Trialgraphix-Kroll Ontrack, New York, USA

Online Publication Date: 01 June 2008

To cite this Article Gjerdingen, Robert O. and Perrott, David(2008)'Scanning the Dial: The Rapid Recognition of Music Genres',Journal of New Music Research,37:2,93 — 100

To link to this Article: DOI: 10.1080/09298210802479268

URL: <http://dx.doi.org/10.1080/09298210802479268>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Scanning the Dial: The Rapid Recognition of Music Genres

Robert O. Gjerdingen¹ and David Perrott²

¹Northwestern University, USA; ²Trialgraphix-Kroll Ontrack, New York, USA

Abstract

Given brief excerpts of commercially recorded music in one of ten broad genres of music, participants in this study were asked to evaluate each excerpt and to assign it to one of the ten genre labels. It was expected that participants would be good at this task, since they were the very consumers for whom much of this music had been created. But the speed at which participants could perform this task, including above-chance categorizations of excerpts as short as 1/4 second, was quite unexpected.

1. Background

Since first presented in 1999 at the annual meeting of the Society for Music Perception and Cognition (SMPC), “Scanning the Dial” has become one of our most requested papers. The fact that it never existed in written form seems not to have dampened its appeal, which may be attributable both to broad interest in the subject matter and to its potential utility in the field of music information retrieval (MIR). Given the relative dearth of human data on the recognition of music genres, the MIR community has made frequent reference to this study, often as a comparison or benchmark with machine classifications of genre (e.g. Tzanetakis & Cook, 2002; Lippens et al., 2004; McKay & Fujinaga, 2006).

One of the original motivations for this study was the first author’s recollection of a question once posed by the American music scholar and composer Charles Seeger (1886–1979). He led a seminar (1961–1971) in the pioneering ethnomusicology programme at the University of California, Los Angeles (UCLA). In the context of music from all over the world played to an international group of students, the question arose, “How long a musical excerpt would one need to hear in order to recognize its culture of origin?” Graduate students in ethnomusicology

devote considerable time and energy to learning the melodic, rhythmic, and textural subtleties of hundreds of diverse musical styles. Perhaps because of this vested interest in expertise, the students guessed that long stretches of music would be required for identification. But the remarkable report from Seeger’s seminar, based on students’ responses to recordings, was that recognition of the source culture was almost immediate. If, for example, the UCLA students had learned the sound of a Javanese gamelan ensemble, they could then recognize it, so to speak, in the blink of an eye. They might need a little more time to distinguish the styles of different regions of Java, but again, much less time than first expected.

Humans have exceptional capabilities when it comes to recognizing the source and significance of sounds. Someone answering a telephone call, for example, can often recognize the caller’s voice almost immediately. Similarly, we can rapidly recognize the characteristic sounds of objects in the environment (e.g. crunching leaves, breaking glass, the patter of rain, the rush of wind). Might music genres, as “objects” in the virtual world of music, lend themselves to similarly rapid “source” identification? Or perhaps might music genres, as proxies for social groups, trigger rapid “friend or foe” responses? Many such questions arise if genre recognition is indeed rapid. No such data, however, had been reported. To get beyond anecdotes it was necessary to determine with some precision how much time ordinary people would require to identify a known music genre.

Shortly before the 1999 SMPC conference, the second author switched from studying cognitive psychology to social psychology. Shortly after the same conference, the first author began working with a Silicon Valley company (Moodlogic, Inc.) whose activities in music recommendation involved proprietary research on music genres. Thus both authors found it difficult to write up a formal version of “Scanning the Dial”. Given the ephemeral, transient

nature of music genres, the original study has now practically passed into the realm of unreproducible results. That is, many music genres have altered their sonic traits in the interim, and college freshman today do not have quite the same ideas of the boundaries between music genres that their predecessors had eight years earlier. Nonetheless, the basic results of “Scanning the Dial” were so striking and consistent across both participants and genres that we feel there is still value in revisiting its methods and results in some detail.

2. Introduction

In the era of vacuum tubes, radios often had a large, glowing, clock-face-like dial. By moving an indicator around the dial the user could change the radio’s tuning in search of a desired station. The phrase “scanning the dial” thus meant passing through various frequencies to find a program or type of program. The dial of a radio has long since ceased to be exclusively circular, and tuning may now be controlled digitally. But the phrase “scanning the dial” still refers to the process of passing through several radio stations in search of desired content. The “scan” function of many contemporary radios automates the process, stopping for perhaps a few seconds at each station. As a feature intended for the mass market, this scan function takes for granted a consumer’s rapid recognition of a radio station’s “format”, which for a music station is another term for the genre of the music that it plays. If, for instance, a consumer is scanning the dial for Hip-Hop, it is safe to say that he or she will not need to spend much time listening to a Country music station to arrive at the conclusion “This is not Hip-Hop”. (Throughout this paper the common names of music genres will be capitalized.) We thus metaphorically extend the phrase “scanning the dial” to the interplay between perceived musical features and human categorizations of meaningfully distinct types of music. We replace the various resonances between radio signals and radio receivers with the cognitive resonances between music heard in the present and the music categorizations developed from music heard in the past.

The advent of digital music available over the internet, something still rather new in 1999, facilitated the collection, control, and presentation of large numbers of music stimuli. In the previous era of vinyl recordings or magnetic tapes, one would have hesitated to envision the presentation of multiple randomized blocks of hundreds of musical excerpts to dozens of participants. The “Scanning the Dial” experiment required the presentation of 20,800 music excerpts. What in the previous era might have necessitated a recording studio, a small army of technicians, and major funding became, in the nascent digital era, a task for one or two personal computers connected to the very large internet collections of excerpts from commercial recordings. These

excerpts had recently been produced with the blessing of the recording industry in order to provide free samples for online shoppers.

While the music industry may attempt to guide and perhaps control the names of music genres, its efforts often lag behind what is happening “on the street”. There is great complexity in the fluid self-organization of the musico-social categories known as genres. Both consumers and the music industry often underestimate this complexity and its dynamics. In popular music, the categorizers (listeners) and the things categorized (songs or “tracks” and their artists) are in constant flux. As a preliminary to the description of empirical work with music genres, it may thus be useful first to list a number of our working hypotheses about this system of both personally and socially constructed categories. Some of these issues have been raised more recently by McKay and Fujinaga (2006), others derive from the general perspectives of musicology, ethnomusicology, and cognitive psychology, and in the case of the first author, practical experience with developing a commercial music-recommendation service (MoodLogic). The frequently contradictory feedback from thousands of customers, each of whom was absolutely convinced of the authority of his or her personal classifications of genre, was a healthy corrective to notions of an essential and fixed definition of any particular genre.

- **Negative Space**

Choosing genre A means concomitantly rejecting genre B and probably most other genres. Genre adjectives with obvious opposites often reflect this relationship. Thus “Cool Jazz” originally meant “not Hot Jazz”, and “Hard Rock” means “not Soft Rock”. Fans of a specific genre can be disturbed by perceived violations of that genre’s negative space. A case in point would be a Jazz fan’s response to the radio format “Smooth Jazz”, whose inclusion of Rhythm & Blues and Adult Contemporary violates the negative space of the traditional genre Jazz.

- **Distinctive Features**

Rightly or wrongly, an individual may associate a specific musical feature with a genre. The use of cues as shortcuts to quick categorizations is widespread in many domains of human cognition. In music, the cue validity of specific instrumental sounds can be quite high. In the 1990s, for instance, New Country and New Traditional Country could often be distinguished by the prominence of solo fiddle and/or steel guitar (+ for New Traditional, – for New Country). Similarly, many traditional Latin genres practically require the percussive sound of claves.

- **The Plasticity Period**

An individual’s personal genre “map” forms during the period from puberty through courtship and early

adulthood. “Baby-boomers” may thus retain their 1960s-era view of music genres well into later life. In the African–American community, those listeners whose plasticity period ended prior to the advent of Hip-Hop now often listen instead to Quiet Storm, a radio format focusing on the Rhythm-&-Blues equivalent of Adult Contemporary. When record labels were a major force in the music industry, their marketing departments directed their campaigns toward consumers in the early stages of acquiring preferred genres and artists. This was the music equivalent of consumer-product companies marketing heavily to young adults who were still choosing their preferred brand of soap or toothpaste. Once a brand is chosen, consumers exhibit significant brand loyalty in later life.

- **The Fisheye-Lens Effect**

An individual recognizes subtle distinctions within the central focus of one or two preferred genres, whereas other genres may be recognized only as stereotypes. A fan of Bluegrass, for instance, knows that Bluegrass comes in many styles that are “easily” distinguishable. Yet that same fan may be certain that all Hip-Hop sounds alike. A fan of Hip-Hop, by contrast, may recognize subtleties in Hip-Hop but only the general sound of all Country genres. In social-psychology terms, this resembles an in-group/out-group effect (Allport, 1954), where we attribute subtlety and other positive traits to our in-group, and have only stereotypical prejudices about out-groups.

- **Demographic Bias**

Age, gender, education, race, class, and other factors may each play a role in mapping social categories onto musical ones. For a young person today, a genre name like “Rhythm & Blues” may represent merely a style of music. Yet that genre name was originally a code word for African–American music produced by “race” labels outside the mainstream of the American music industry. Gender can also play a major role in genre preference. In the 1990s the audiences for Lilith Fair (female) and Thrash Metal (male) were so obviously gender specific as to make people stare at a fan of the “wrong” gender. The (in)famous “no-two-female-artists-in-a-row” rule on Rock radio stations both responded to and helped refine gender bias within that genre.

- **The Uniqueness Paradox**

Almost everyone asserts that his or her musical tastes are unique, yet the music industry largely serves mass tastes. If every person’s tastes were unique, how could Britney Spears have sold eighty-three million CDs? As with young listener’s ideas about musical

“authenticity”, consumer participation in the commercial music marketplace may involve a measure of “suspension of disbelief”.

- **“Natural” Categories**

Genres with short, simple names tend to be the core categories around which cluster other, secondary categories. Thus “Rock”, “Jazz”, or “Country” are core genres whereas “New Traditional Country” or “Cool Jazz” or “Chamber Rock” would be considered sub-genres. In the main, the naming patterns of genres conform to norms for how ordinary people devise categories to describe everyday objects or experiences (see Rosch, 1973).

- **Genre Anchors and Overlaps**

Listeners can ascribe a song to multiple genres through a type of triangulation from known positions. “Country, leaning toward Blues-Rock with a touch of Rockabilly” is, for listeners knowledgeable of those genres, a fully intelligible description. Part of this ability and common mode of description may stem from the fact that recordings are collectively produced artworks, with each member of a band having slightly different musical roots, preferences, and modes of performance. Producers, record executives, and lead artists can further shape the collective artwork. In recent decades, mixing known styles in popular music has been a common practice. The areas where genres overlap, both in musical style and in consumer preferences, are highly contested in the marketplace. The industry responses to real or potential overlaps are expressed through shifts in radio formats or in attempts at the subtle re-branding of artists.

- **The Customer is Always Right**

If person X says that a song is in genre B, person X is subjectively correct regardless of what anyone else says. Only in reference to group norms can one properly talk about the “correct” genre. An individual’s private mapping of the space of music genres is a datum. To speak of a social group’s mapping is to make a generalization from an inferred set of data.

- **Exemplars Come First**

Experienced listeners may develop rich categorical representations of genres; inexperienced listeners may associate a genre with one or two known examples or artists. At the time of this study, Latin meant only “Ricky Martin” to many of the participants, while a few could distinguish New York Salsa from Cuban Mambo. Categorizations may be the result of the brain simultaneously recalling a host of related exemplars (cf. Nosofsky et al., 1989).

- **We Hear Through the Sound to its Meaning**

When we listen to speech from a familiar voice, we attend to the message and not to an agglomeration of phonetic details. Similarly, listeners experienced in a music genre attend to the communication of combined musico-verbal messages and not to clusters of individually noticed musical features. For many listeners this process extends further to the reception of an emotional message from the virtual persona of an artist. When the artist is strongly associated with one genre (e.g. Dolly Parton with Country), that association may trump musical distinctions that would otherwise indicate a different genre. Genres in which an artist persona is unclear or unspecified (e.g. Electronica or other machine-made musics) may at times confound predictions of perceived genre.

3. Methods

3.1 Genres

We chose ten main genres of popular music widely recognized in the late 1990s. They were (listed alphabetically):

1. Blues
2. Classical
3. Country
4. Dance
5. Jazz
6. Latin
7. Pop
8. R&B
9. Rap
10. Rock

In the intervening years some of these categories have lost their prominence or changed their meanings. Dance and Pop have largely dissolved into other genres, R&B (“Rhythm and Blues”) and Rap have partly merged into Hip-Hop, Latin has diminished in the mass “Anglo” market post Ricky Martin, and Rock has continued to splinter into various niches. Nevertheless, at the time of this study, those ten genres were prominently featured in record stores, by online merchandisers, on fan-sites, and in the responses to a questionnaire filled out by our participants. Given a temporal lag in culture-wide main-genre categorizations, the ten genres chosen thus probably represented a white-American consensus of the mid 1990s. In particular they represent the American music industry before it fully recognized the importance of Hip-Hop and before the concomitant decline of truly mass-market stars like Michael Jackson or Cher made the meaning of Pop more tenuous.

3.2 Stimuli

Pilot testing demonstrated that our original guess at the appropriate minimum lengths of recorded excerpts (1, 2, 3, 4 and 5 s) was wrong by almost an order of magnitude. Participants were so adept at recognizing genre that their performance across those durations appeared subject to a strong ceiling effect at perhaps 2 s. The durations thus selected for the experiment were 250, 325, 400, 475, and

3000 ms, with the 3000 ms duration included to quantify a ceiling for comparison with the brief excerpts.

We downloaded eight tracks in MP3 format for each genre. We accepted the genre classifications of the leading web-based music vendors of that time (CDnow, BMG, Tower Records), avoiding any songs that might suggest genre crossover. For each group of eight tracks, four excerpts were chosen to have prominent “vocals” (at least one singer) and four were “instrumental” (no voice during the excerpt). For each track, we selected a 3000 ms excerpt that we subjectively felt was characteristic of both the song and its genre. Given the working methods of the music industry, that often meant excerpting the “hook” or characteristic refrain of the song. From the 3000 ms excerpt, four small excerpts were taken, corresponding to the above-mentioned durations of 250, 325, 400, and 475 ms.

3.3 Participants

Fifty-two Northwestern University students who were enrolled in a first-year psychology course volunteered for this study. The thirty-one females and twenty-one males had a mean age of 19.5 yr, the range being 18.25 to 22.25. They reported listening to music 24.2 h per week, though a few reported listening to as much as 100 h per week (thus a high standard deviation of 24.4 but a median of only 17). Although Northwestern has a large School of Music, the participants in this study would not generally be considered musicians (no one was a music major, 80% had no aural training, 75% had no music-theory training, 77% had never sung or played in an ensemble, and 33% had never had lessons for voice or instrument). “Ordinary undergraduate fans of music” might be an apt descriptor, and that would make them one of the target audiences of the music industry. In other words, they were the “receivers” for whom much of the content in the music “signal” was optimized.

3.4 Protocol

Participants were tested individually on an IBM-compatible PC. They were given ten practice trials (one excerpt from each style), with the instructions to indicate, via labelled keys corresponding to each style, which style of music they thought the excerpt best represented. The exemplars were then presented in five blocks of 80 trials. The first four blocks contained randomly presented 250, 325, 400 and 475 ms excerpts. The order of presentation of the first four blocks was randomized between subjects so as to distribute any order effects. The final block contained the randomly ordered 3000 ms excerpts. This was done so that participants’ responses to the excerpts at shorter intervals could be compared to their responses to the longer excerpts (corresponding to their subjective, unhurried classification of the piece of music), as well as to the music companies’

classification of the excerpts. A brief pause was allowed between testing blocks.

Each participant listened to and categorized 400 excerpts: 5 (durations) \times 10 (genres) \times 2 (voice/instrument) \times 4 (excerpts).

4. Results

At ceiling performance (3000 ms) participants agreed with the genres assigned by music companies about 70% of the time. This does not mean that participants were “wrong” 30% of time. As mentioned, “correct” in the context of music genres involves some subjectivity. Even different segments of the music industry can disagree about the genre of a particular song or track. In the 1990s, for instance, many people conceived of Dance as pop-oriented, cheerful songs, others thought of Dance as Hip-Hop-oriented urban tracks, and still others were convinced that Dance meant Electronica of the type played at raves. Each person was “correct” (The Customer Is Always Right), but the joint statistics of such individuals may easily confound failures of recognition with confidently recognized but alternative categorizations.

If the ultimate meaning of each genre name is a question for qualitative humanistic or sociological discourse, the degree to which participants were self-consistent is an empirical question amenable to quantitative analysis. Figure 1 shows the extent to which, for short excerpts from a given genre, participants agreed

with their individual responses to 3000 ms excerpts from the same song (the short excerpts were segments of the long excerpts; the 3000 ms excerpts were heard only at the end of the experiment).

There is considerable variation across genres. Given the ten genres presented, the chance level for any response would be 10%. Averaged over time intervals 250 to 475 ms, for every genre responses far exceeded chance ($t(52) \geq 10.58$, $p < 0.001$, for each t -test for each genre). Even for just the 250 ms excerpts, guesses were greater than chance for every genre ($t(52) \geq 3.01$, $p < 0.001$, for each t -test for each genre).

Figure 2 shows this categorization data by time interval, and further broken down according to instrumental and vocal excerpts. Simple within-subjects comparisons of mean proportions for each time interval collapsed across genre revealed that there was a main effect of duration: classification (i.e. agreement with the 3000 ms response) improved as the time interval increased from 250 to 325 ms ($M_d = 0.08$, $SD = 0.07$, $t(52) = 8.41$, $p < 0.001$), and from 325 to 400 ms ($M_d = 0.04$, $SD = 0.06$, $t(52) = 5.11$, $p < 0.001$). The improvement from 400 to 475 ms, however, was not significant ($M_d = 0.01$, $SD = 0.05$, $t(52) = 0.90$, $p = 0.37$).

Figure 2 also shows that there was a main effect of voice/no-voice: classification was more accurate at each time interval for instruments-only exemplars ($M_d \geq 0.04$, $t(52) \geq 2.60$, $p = 0.01$), except at 250 ms where there was no difference ($M_d = 0.01$, $t(52) = 0.73$, $p = 0.47$). The mean percentage of participants agreeing with their 3000 ms classification, averaged over the 250, 325, 400,

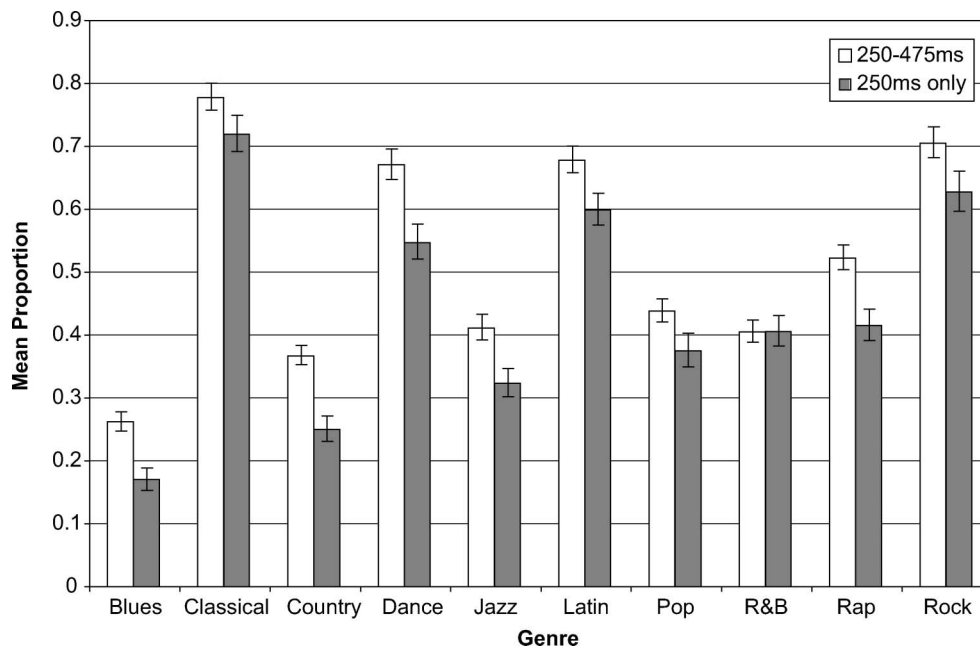


Fig. 1. Mean proportion of 250–475 ms categorizations matching 3 s self-categorization (collapsed over instrumental and vocal excerpts).

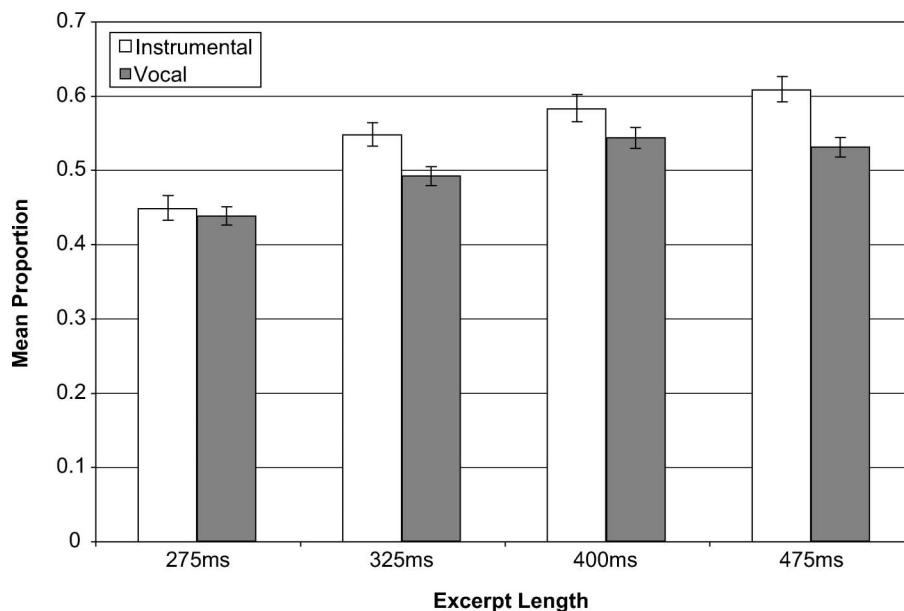


Fig. 2. Mean proportion of instrumental versus vocal 250–475 ms categorizations matching 3 s self-categorization (collapsed over genre).

and 475 ms time intervals, was 54.62% for instrumental excerpts, versus 50.04% for vocal excerpts.

In a pilot study with a small group of students majoring in music theory, the corresponding “combined” percentage at 250 ms was 43%, essentially the same as the 44% for the non-musicians in the main study. This suggests little or no effect of formal music training on this task.

5. Discussion

It takes about 250 ms or more to say the word word. In such a short interval a popular song may present only one harmony, perhaps only a tone or two in the bass, and perhaps only a tone or two in the melody. At a moderate tempo of 110 beats per minute, less than half a beat will fit within 250 ms. By almost any interpretation of music theory, this time interval is too short to allow for a classification of genre. For example, knowing that the melody was G4 then E4, the bass was C3, the harmony was a major chord, and a drum sound occurred 100 ms into the excerpt would provide no way to separate Country from Pop or Rock. Yet the participants in this study accomplished this and similar tasks with relative ease. How did they do it?

What is called, for lack of a better term, “timbre” encompasses all the spectral and rapid time-domain variability in the acoustic signal. Such information can be highly indicative of particular genres. There is a perceived “twang” to Country, for instance, and the particular timbre of the 12-string guitar was once

emblematic of the genre Folk in the period 1955–1965. Similarly, enormous acoustic energy in extreme low frequencies would point to Rap and later Hip-Hop, while at the same time pointing away from the older sub-genres of Country. The formalized, “strict” tone production and articulation of Classical contrasts sharply with the “looser” values in popular styles. The results of this study suggest that a highly reduced combination of melodic, bass, harmonic, and rhythmic features can help to classify genre if these features are coupled with an accurate acoustic signal. This study does not, however, shed light on the relative weighting of music-theoretical and acoustic features in the categorization results of the participants. The original title of this study, as submitted to the 1999 SMPC conference, was “Scanning the dial: An exploration of factors in the identification of musical style”. We had initially assumed that music-theoretic factors would emerge, but the participants’ success at identifying genre at unexpectedly brief time intervals made such a factor analysis untenable or at least beyond our expertise.

Timbre may also be a factor in the slightly but consistently superior classification of instrumental (non-vocal) versus vocal excerpts. It is possible that timbral variation across some voices is less indicative of style than across non-vocal instruments. It is also possible that the voice, given special prominence by commercial recording engineers, has a masking effect that made the timbral cues of other instruments more difficult to discern.

To evaluate the extreme cases – the 70% agreement for 250 ms of Classical and the only 27% response for the

average 250–475 ms excerpts of Blues, it may be helpful to think of an abstract space of genres. In that space, the location of Classical might be off by itself, a considerable distance from any popular genre. The location of Blues, by contrast, might be overlapped by sub-genres of Rock, R&B, Jazz, Pop, and so forth. So when presented with a very short excerpt, a participant may have had difficulty discriminating between these occasionally closely related, sonically similar genres. There may also be response biases due to relative unfamiliarity, with Blues being directed less toward a mass market than, say, R&B, and more toward an audience older than the participants in our study.

To give a concrete example of how genre ascription can shift with reduced information, a 3 s excerpt of one exemplar classified as Dance by the music industry was classified by 42 out of 53 respondents as Dance, a 79% agreement with industry norms. This is typical of our overall results and indicative of a ceiling in the 2–3 s range (i.e. respondents in a pilot study were only marginally more in agreement with the music industry if they listened to an entire track). Yet 21 of the same respondents (40%) classified the shorter excerpts of that same exemplar as either R&B or Pop. Because categorizing an excerpt as Dance includes inferring suitability for entraining body movements, it is possible that the short excerpts could not afford a confident categorization along those lines and that the respondents thus defaulted to the broader and possibly super-ordinate categories of R&B or Pop.

We did not notice a strong effect of gender in our data. Moreover, since two-thirds of our participants were female, our data set may not be ideal for exploring the effect of gender in these categorizations. The music industry certainly recognizes gender as an important factor, and gender may be one type of Demographic Bias in genre categorization. One possibility is that gender has a relatively weak effect that does not emerge with such short musical excerpts. Another possibility is that gender has an effect in music selection, but not recognition. Yet another possibility is that gender might effect recognition of sub-genres but not main genres.

In the period 2000–2002 the MoodLogic company did extensive online surveys where people rated popular songs on various perceptual axes. One of these was a Likert scale of seven choices ranging from “Weak Bass” to “Strong Bass”. Figure 3 shows contrast-enhanced, normalized mean responses to every song rated by approximately 4000 listeners who had submitted demographic profiles. The x axis represents increasing categories of age (from <14 to >55), the y axis represents increasing evaluations from “Weak Bass” (negative numbers) to “Strong Bass” (positive numbers):

The graph indicates that, for the same universe of songs, older evaluators perceived the bass to be stronger than younger evaluators. Notice another Demographic

Bias in the histograms of Figures 4 and 5. Figure 4 shows responses on the “Peaceful . . . Aggressive” scale from more than 260 evaluators who described their occupation as “Homemaker”.

Figure 5 reflects responses to the same songs on the same perceptual axis, but by the more than 600 evaluators who described their occupation as “Comp/Tech/Engineering”.

“Homemakers”, who were overwhelmingly female, evaluated the same repertoire of songs as being more aggressive than the “comp/tech/engineers”, the great majority of whom were male. These evaluation biases, which pose classic nature-versus-nurture questions, may or may not impinge on genre classifications. The present

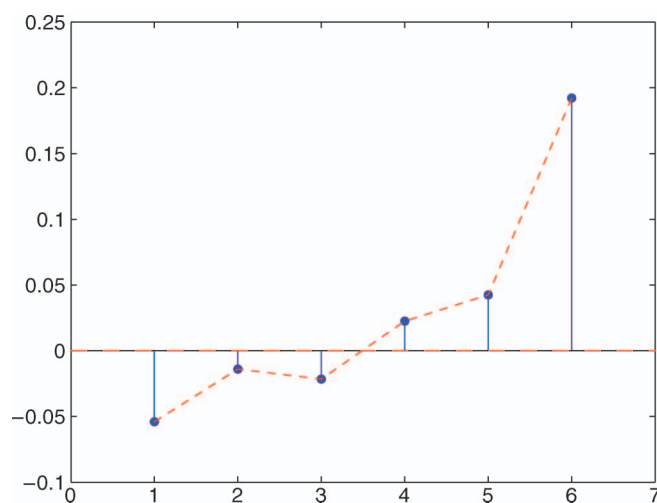


Fig. 3. Increasing categories of age (x -axis) plotted against the perception of the bass being strong (y -axis).

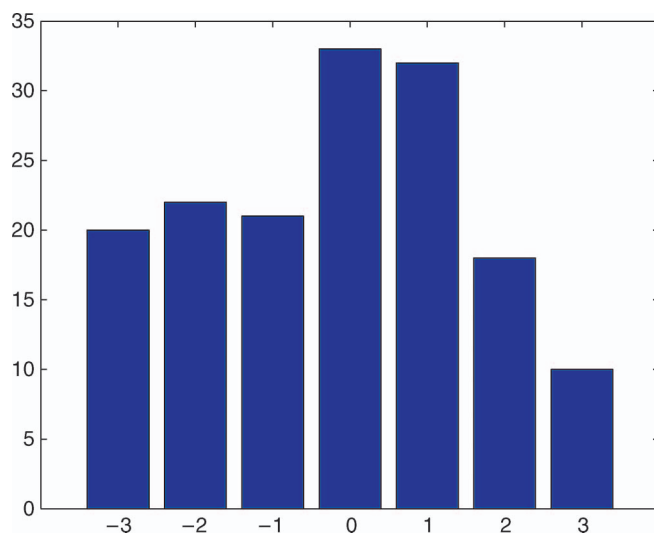


Fig. 4. Overall ratings of a large repertoire of songs on the scale “Peaceful . . . Aggressive” (“very peaceful” = -3 , “very aggressive” = $+3$) by self-described homemakers.

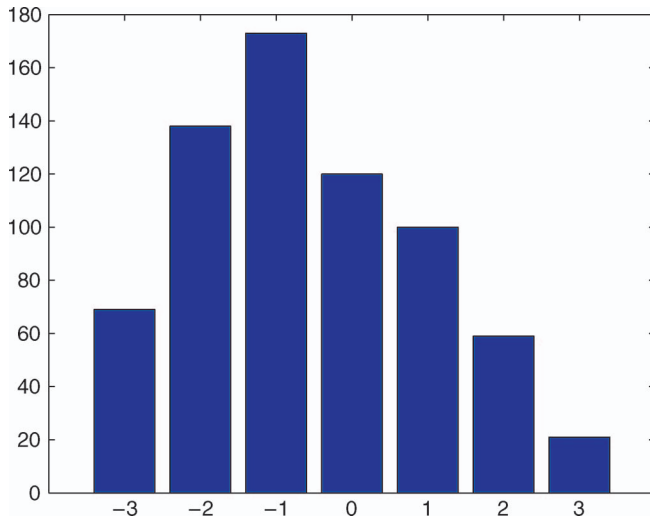


Fig. 5. Overall ratings of a large repertoire of songs on the scale “Peaceful ... Aggressive” (“very peaceful” = -3, “very aggressive” = +3) by self-described workers in “Comp/Tech/Engineering” fields.

study, with its focus on rapid categorizations, cannot speak to these larger questions. But the Demographic Bias in these data suggest that for machine categorizations of music, especially those connected with recommendation services, a prior categorization of the listener may improve predictions of the listener’s categorizations of music genre. In other words, there may be a space of listeners, a space of music genres, and a translation matrix that relates the one to the other.

6. Conclusion

This study demonstrates the near immediacy of genre identification as a response to a musical stimulus. If genre identification, as a contextualizing categorization,

required the prior classifications of component features like melody, bass, harmony, and rhythm, then it is unlikely that such rapid identification would have been possible. That is, from a single tone one could not infer any reliable information about melody or rhythm. Rather, it seems highly probable that the rapid recognition of musical genre occurs concomitantly with the decoding of component features. In a manner reminiscent of Gestalt effects, it would appear that listeners can achieve a global categorization of genre at least as fast as they can categorize any component feature. Presumably genre categorization is a learned response, and the rapidity of that response may reflect the great social and personal significance of music genre in the modern world.

References

- Allport, G.W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Lippens, S., Martens, J.P. & De Mulder, T. (2004). A comparison of human and automatic musical genre classification. In *Acoustics, Speech, and Signal Processing, 2004, Proceedings (ICASSP'04). IEEE International Conference Proceedings, 4*, 233–236.
- McKay, C. & Fujinaga, I. (2006). Musical genre classification: is it worth pursuing and how can it be improved? In *Proceedings of the International Conference on Music Information Retrieval* (pp. 101–106). Victoria, BC.
- Nosofsky, R.M., Clark, S.E. & Shin, H.J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2*, 282–304.
- Rosch, E.H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.
- Tzanetakis, G. & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*, 293–302.