Journal of Experimental Psychology: Human Perception and Performance

Patterns of Individual Differences in the Perception of Missing-Fundamental Tones

D. Robert Ladd, Rory Turnbull, Charlotte Browne, Catherine Caldwell-Harris, Lesya Ganushchak, Kate Swoboda, Verity Woodfield, and Dan Dediu

Online First Publication, February 11, 2013. doi: 10.1037/a0031261

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/13/$12.00 DOI: 10.1037/a0031261

Patterns of Individual Differences in the Perception of Missing-Fundamental Tones

D. Robert Ladd, Rory Turnbull,
and Charlotte Browne
University of Edinburgh

Catherine Caldwell-Harris
Boston University

Lesya Ganushchak
Max Planck Institute for Psycholinguistics, Nijmegen,
The Netherlands

Kate Swoboda and Verity Woodfield
University of Edinburgh

Dan Dediu
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Recent experimental findings suggest stable individual differences in the perception of auditory stimuli lacking energy at the fundamental frequency (F0), here called missing fundamental (MF) tones. Specifically, some individuals readily identify the pitch of such tones with the missing F0 ("F0 listeners"), and some base their judgment on the frequency of the partials that make up the tones ("spectral listeners"). However, the diversity of goals and methods in recent research makes it difficult to draw clear conclusions about individual differences. The first purpose of this article is to discuss the influence of methodological choices on listeners' responses. The second goal is to report findings on individual differences in our own studies of the MF phenomenon. In several experiments, participants judged the direction of pitch change in stimuli composed of two MF tones, constructed so as to reveal whether the pitch percept was based on the MF or the partials. The reported difference between F0 listeners and spectral listeners was replicated, but other stable patterns of responses were also observed. Test-retest reliability is high. We conclude that there are genuine, stable individual differences underlying the diverse findings, but also that there are more than two general types of listeners, and that stimulus variables strongly affect some listeners' responses. This suggests that it is generally misleading to classify individuals as "F0 listeners" or "spectral listeners." It may be more accurate to speak of two modes of perception ("F0 listening" and "spectral listening"), both of which are available to many listeners. The individual differences lie in what conditions the choice between the two modes.

Keywords: missing fundamental, pitch perception, individual differences

Supplemental materials: http://dx.doi.org/10.1037/a0031261.supp

A missing fundamental (MF) tone is an artificially constructed acoustic stimulus consisting of a number of component frequencies, chosen so that they could be the harmonics of some fundamental frequency (F0) that is itself not present in the stimulus. For example, consider a tone consisting of energy at 750 Hz, 1000 Hz, and 1250 Hz. The lowest common factor of these frequencies is

D. Robert Ladd, Rory Turnbull, and Charlotte Browne, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, Scotland; Catherine Caldwell-Harris, Department of Psychology, Boston University; Lesya Ganushchak, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Kate Swoboda, School of Philosophy, Psychology and Language Sciences, University of Edinburgh; Verity Woodfield, School of Philosophy, Psychology and Language Sciences, University of Edinburgh; Dan Dediu, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Rory Turnbull is now at Department of Linguistics, Ohio State University, and Kate Swoboda is now at the School of Life and Health Sciences, Aston University, Birmingham, England.

Thanks to Eddie Dubourg and Simon Kirby for their technical contributions and to Richard Shillcock, Morten Christiansen, Tim Bates, and Antje Meyer for discussion. The pilot experiments, which determined the general direction of subsequent work, were carried out jointly by D. Robert Ladd, Rory Turnbull, and Dan Dediu. The other authors were each involved in one of the four larger studies: Catherine Caldwell-Harris in Exp. 3b, Lesya Ganushchak in Exp. 4a, Kate Swoboda in Exp. 4b, and Charlotte Browne and Verity Woodfield in Exp. 4c. Exp. 4c was part of two student papers. D. Robert Ladd prepared the stimuli for all experiments and had primary responsibility for writing the paper, and Dan Dediu had primary responsibility for the statistical analyses.

Correspondence concerning this article should be addressed to D. Robert Ladd, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh EH8 9AD, Scotland. E-mail: bob.ladd@ed.ac.uk

250 Hz and, in general, such a tone is often perceived as having a pitch of 250 Hz; that is, the pitch percept may be based on a frequency that is, in some sense, not physically present in the stimulus. This frequency—also referred to in the literature as "virtual pitch" (e.g., Terhardt, 1979), "periodicity pitch" (e.g., Licklider, 1951), and "residue pitch" (e.g., Schouten, 1940)—is the missing fundamental. However, it is also possible to perceive the MF tone just described as a chord consisting of the component frequencies (the "partials") that are actually present in the stimulus—specifically, in musical terms, as an inverted major triad (roughly a very flat G_5 C_6 E_6 [g″ c‴ e‴]). The starting point for this article is the finding that many individuals seem to have stable biases in the way they perceive MF stimuli, preferentially hearing the pitch of the stimulus either on the basis of the MF or of the partials that are actually present.

The source of these individual differences is not known. Recent interest in this topic has arisen within cognitive neuroscience, especially among those interested in music perception and cognition. Some of this work seeks to correlate different patterns of responses to MF stimuli with neuroanatomical (e.g., Schneider et al., 2005) or neurophysiological (e.g., Patel & Balaban, 2001) differences; other work emphasizes the influence of experience, particularly musical training, on the patterns of perceptual responses (e.g., Seither-Preisler et al., 2007). However, it is also known that there are purely physical effects that influence the actual acoustic nature of signals consisting of a small number of partials, and that these may affect the cochlear response to the signals; probably the most important effect of this sort is the existence of "combination tones" (see, e.g., Terhardt, 1974; Moore, 2012). There is a separate line of recent research on MF perception among hearing researchers that seeks to understand these basic physical mechanisms (e.g., Bernstein & Oxenham, 2006; Gockel, Plack, & Carlyon, 2005; Gockel, Carlyon, & Plack, 2010; see Moore & Gockel, 2011, for a recent review). It is entirely possible that some of the individual differences under discussion here are based on different cochlear responses to differences in the signal, rather than originating in the brain.

However, the present article is concerned not with the basis of the behavioral differences but with a clearer definition of the differences themselves. Recent work is extremely diverse methodologically and has focused on testing hypotheses about the effect of specific individual differences (e.g., differences of musical training) on the perception of MF stimuli. Moreover, it has tended to proceed as if the behavioral differences are straightforwardly binary, describing individuals as belonging to one of two basic types of listeners. Our investigations have shown that this approach oversimplifies the nature of the individual differences, and we believe that this oversimplification directly affects our ability to look for their underlying causes. Our aim in this article is to present a more refined characterization of the behavioral differences, which will be of use to subsequent research on any aspect of the MF phenomenon.

The MF Task

Basic Design

The first systematic exploration of individual differences in responses to MF tones was carried out by Smoorenburg (1970), who seems to have stumbled on the existence of the individual differences while researching the basic physics of the phenomenon (p. 927). Smoorenburg developed an ostensibly simple way to determine whether a listener is taking the missing F0 or one of the partials as the pitch of a MF tone. By presenting MF tones in pairs, he was able to construct stimuli that would appear to go either up or down in pitch from the first member of the pair to the second, depending on whether the pitch of the individual members of the pair was being perceived on the basis of the MF or of the partials. This behavioral task is the experimental tool on which subsequent research has been based.

The basic design of stimuli in the MF task is diagrammed in Figure 1. In this example, it can be seen that the MF "goes down" (i.e., is lower in Tone B than in Tone A), but the lowest partial actually present in the stimuli "goes up" (i.e., is lower in Tone A than in Tone B). This ambiguity can be achieved even while keeping the highest partials at the same frequency in both tones; all that is needed is to treat that top frequency as the nth harmonic in Tone A and the $(n+1)$th harmonic in Tone B. To avoid misunderstanding, it is worth mentioning that the terms "Tone A" and "Tone B" are used only for clarity of reference and imply nothing about order of presentation. In the various studies discussed here, actual stimulus pairs were of course presented in either order (AB or BA) or in both orders. No source reports any order effects, but as we shall see, such effects do occur, which complicates the interpretation of what listeners are actually doing in the MF task.

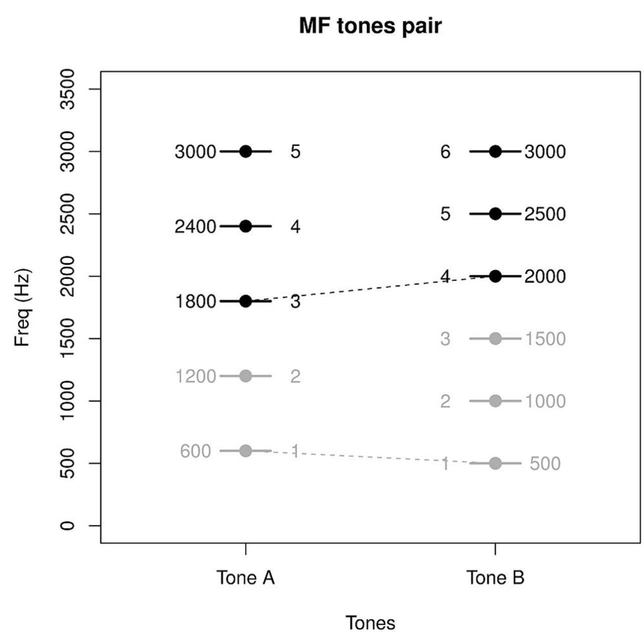

Figure 1. Basic design of missing fundamental (MF) task stimuli. Tone A (on the left) consists of three partials that could be the 3rd, 4th, and 5th harmonic of a fundamental frequency (the "first harmonic") that is not physically present in the signal. Tone B (on the right) also consists of three partials, which could be the 4th, 5th, and 6th harmonics of a fundamental frequency (also not physically present). Crucially, the MF in Tone B is lower than that in Tone A, and the lowest frequency actually present in Tone B is higher than the lowest frequency actually present in Tone A.

Individual Differences in MF Perception

Smoorenburg's (1970) experiment suggested that most individuals fairly consistently perceive the pitch of the MF tones either in terms of the missing F0 or on the basis of the component frequencies. His data also made it appear that there are roughly equal numbers of the two types of listeners. However, his procedure involved only two different stimuli, presented repeatedly (i.e., a single pair of "Tone A" and "Tone B" in both orders of presentation). If there is a genuine source of individual difference, it is not surprising that this procedure would lead to a strong separation of the two response patterns. More recent studies seem to show that if listeners are presented with a range of different MF stimuli, their behavior may be more variable, and that the properties of the stimulus may have consistent influences on which way listeners tend to hear it.[1] We focus here on a comparison of two large studies—Schneider et al. (2005) and Seither-Preisler et al. (2007).

Schneider et al. (2005) was a large study of musicians and nonmusicians, with the primary aim of relating differences in MF perception to differences in neuroanatomy, specifically to differences in the volume of the pitch-detection areas in left and right Heschl's gyrus. A secondary aim was to explore the effect of certain stimulus variables (e.g., number of partials present in the stimulus tones) on the perception of MF tones. Schneider et al. reduced listeners' overall pattern of responses to a quotient whose value ranges from -1 to $+1$, according to the proportion of responses based on F0 and on the partials. They report a bimodal (broadly "U"-shaped) distribution in the value of this quotient, with a minimum in the middle of the range (around 0, where an individual's responses are mixed). On this basis, they divide the range in half and classify listeners as "F0 listeners" or "spectral listeners." We will adopt this terminology here.[2] Schneider et al. also report that, on average, spectral listeners have greater cortical volume in Heschl's gyrus in the right hemisphere than in the left, whereas F0 listeners have greater volume in the left than in the right. They found no consistent difference in responses or in hemispheric asymmetry between musicians and nonmusicians, but report overall larger Heschl's gyrus volume in musicians. Related work by Schneider and Wengenroth (2009) suggests that there may be differences among musicians depending on their instrument or the type of music they play; for example, jazz musicians are more likely to be spectral listeners than classical musicians.

Seither-Preisler et al. (2007) also studied musicians and nonmusicians; they did not do any brain imaging, but their hypotheses are implicitly driven by assumptions about brain plasticity, specifically the effect of musical training. Like Schneider et al., their materials manipulated a number of different stimulus variables, but the variables they explored differed quite considerably from those studied by Schneider et al. They also used very different (and more complex) statistical reductions of individuals' behavioral response patterns that ultimately abstracted away from the effect of stimulus variables. Like Schneider et al., they found that many participants responded as F0 listeners or spectral listeners, and indeed, they report a sharper dichotomy between the two groups than was found by Schneider et al. However, this sharper dichotomy is due in part to their analysis procedures, which led them to exclude roughly a quarter of their participants on the grounds that their responses were not reliably distinguishable from guesswork. They also, unlike Schneider et al., showed a clear effect of musical training, with professional musicians responding far more often as F0 listeners. Note in this connection that Seither-Preisler et al.'s repeated references to "guessing" may seem to suggest that there is a right answer (viz., F0 response), an implication that we find unjustified.

Stimulus Variables in the MF Task

One of the striking features of the two studies just summarized is that they make very different methodological choices in their procedures and in constructing their stimuli, yet both find evidence for Smoorenburg's (1970) basic conclusion that listeners exhibit two essentially different types of behavior in processing MF stimuli. Other recent studies, based on still other methodological approaches, lead to the same conclusion. For example, Patel and Balaban (2001), a study of neural activity in pitch perception with a focus on the relation between time-domain and frequency-domain processing, also finds clear evidence that individuals tend to favor one of two different modes of behavior. The fact that these differences show up in a wide variety of experimental situations suggests that the underlying phenomenon is very robust.

At the same time, early psychoacoustic work into the nature of MF perception in general (Plomp, 1967; Ritsma, 1962, 1963a, 1963b) has demonstrated that stimulus properties can have consistent effects on listeners' responses. Systematic manipulation of stimulus variables in subsequent work (e.g., Moore, Glasberg, & Peters, 1985; Houtsma & Fleuren, 1991) further established the role of stimulus properties in determining response patterns, independent of individual differences. These effects were not absent from Schneider et al. and Seither-Preisler et al.'s results. Two such findings emerge clearly from these two articles:

- As the musical interval between the missing fundamentals in Tone A and Tone B increases, listeners are more likely to give F0 responses. This effect was demonstrated clearly by Seither-Preisler et al. (p. 746, Figure 3; cf. Meddis & Hewitt, 1991; Moore et al., 1985).
- As the number of partials in the tones increases, listeners are also more likely to base their pitch judgment on the missing F0. This effect was systematically shown by Schneider et al. (p. 1242, Figure 1d; cf. Faulkner, 1985; Ritsma, 1962).

This means that, irrespective of an individual's bias toward F0 or spectral listening, responses can be influenced by differences of detail in the stimuli. It therefore seems important to consider methodological choices in stimulus construction more closely. Unfortunately, this is not as straightforward as it might sound, because the stimulus variables are highly interdependent. We cannot simply vary them orthogonally to explore their effects.

[1] Louis Pols (personal communication, September 2011) tells us that he worked in the same lab as Smoorenburg at the time of the experiments on which the 1970 paper was based, and says that Smoorenburg was well aware that some MF tones would elicit F0 percepts from most listeners. Stimuli had to be carefully chosen in order to draw out the difference between individuals.

[2] "Synthetic" and "analytic" are two common terms used for F0 and spectral listeners, respectively, and are widely used in the literature (e.g., Schneider & Wengenroth, 2009). Although this pair of terms has a long history (Houtsma & Fleuren, 1991, attribute the terms to Hermann von Helmholtz), we prefer the terms from Schneider et al. (2005), which are more theoretically neutral.

This interdependency can be illustrated clearly by the relation among what we might refer to as top frequency (the frequency of the highest partial), harmonic rank (the position of the partials in the harmonic series, e.g., 5th and 6th harmonics), and the interval between the missing F0 of Tone A and Tone B. If top frequency is held constant within a stimulus pair (as was done by Schneider et al.), then interval is completely determined by the choice of harmonic rank for the two stimulus tones (or vice versa); if interval is systematically varied (as was done by Seither-Preisler et al.), then the top frequency of the two stimulus tones is completely determined by their harmonic rank (or vice versa). For example, if the top frequency of a stimulus pair is kept constant at 600 Hz and we specify the top partials in tones A and B as having harmonic rank 5 and 6, respectively—which roughly corresponds to the procedure of Schneider et al.—the interval will necessarily be a minor third (three semitones), because the ratio of the virtual F0 of the two stimulus tones will be 6:5 (120 Hz and 100 Hz). If the top frequency is held constant at 600 Hz and we want to specify an interval of a fifth (ratio 3:2), we would have to use harmonic rank 4 and 6 (or 6 and 9, or 8 and 12, etc.). Conversely, if we specify an interval of a fifth and also specify the harmonic rank of tone A and B—which corresponds roughly to the procedure of Seither-Preisler et al.—then the top frequency of one stimulus tone will necessarily be higher than the other. Similar interdependencies affect other stimulus variables; fuller discussion is beyond the scope of this report.

This interdependency makes it difficult to interpret some of the findings reported in the articles under consideration or to investigate apparent contradictions. The most obvious discrepancy here involves overall frequency level and harmonic rank. Schneider et al. report an effect of "average spectral frequency" (p. 1242, Figure 1c): As the average frequency of the stimulus tones increases, so, too (albeit rather irregularly), does the number of F0 responses. At the same time, they also report an effect of harmonic rank, such that partials *lower* in the harmonic series evoke more F0 responses (p. 1242, Figure 1d); Seither-Preisler et al. (p. 745 ff.) mention a similar effect of harmonic rank in a variable they call "spectral profile." These findings make exactly opposite predictions about the effect of manipulating the partials in a MF tone pair at a given F0 level: Higher partials will raise the average spectral frequency and therefore should lead to more F0 responses, yet higher partials will also be higher in the harmonic series and therefore should lead to more spectral responses. Furthermore, in a pair of MF tones constructed according to Schneider et al.'s procedures, higher partials will yield smaller intervals between the missing F0 of the two tones, which (given Seither-Preisler et al.'s results) should lead to more spectral responses as well. Because it is physically impossible to vary harmonic rank, MF interval, and average spectral frequency orthogonally while keeping F0 within a constrained range, we cannot resolve these contradictory predictions in conventional experimental ways.

Classification of Listeners

Given the forced-choice approach of the experiments just discussed, labels such as "F0 listening" and "spectral listening" can certainly be applied to individual *responses*. However, it is less clear that these labels can also be appropriately used to describe the overall behavior of *listeners*—that is, whether individuals clearly fall into two groups with distinct behavioral strategies. It seems likely that there really are distinct behavioral strategies, but the matter is not simple and it depends, to some extent, on how we quantify overall patterns of individual responses.

Schneider et al. add each participant's responses together and compute an individual "index" that expresses the proportion of F0 and spectral responses on a scale from -1 to $+ 1$. We refer to this score in what follows as the Schneider index (SI). Their formula is as follows:

$$SI = \frac{sp - f0}{sp + f0} \tag{1}$$

where *f0* refers to the number of F0 responses and *sp* refers to the number of spectral responses. Seither-Preisler et al. use a similar score to describe individual performance on their Auditory Ambiguity Test (AAT), which simply reports the overall proportion of F0 responses on a scale from 0 to 1.0. These two measures are completely equivalent, with an SI of -1 corresponding to 1.0 on the AAT, an SI of $+1$ corresponding to 0, and an SI of 0 corresponding to 0.5.[3] As previously noted, both teams report bimodal distributions of these quantitative measures, with many listeners having scores near the ends of the range and fewer in the middle.

The most important problem with this approach to data reduction is intraindividual consistency. Some participants give completely consistent responses—that is, 100% of their responses are either "F0" or "spectral." In these cases, there is no issue about describing individuals as "F0 listeners" or "spectral listeners." However, many participants give a mix of responses, which can yield an SI near 0. It is not immediately obvious how to treat such mixed behavior.

Schneider et al. hypothesized that some degree of inconsistency might arise through what they called "octave shifting," that is, perceiving the *second* harmonic (one octave higher than the missing F0) as the pitch of a MF tone. They attempted to allow for this kind of inconsistency by including control stimuli in which Tone A actually includes the F0 (in terms of the example shown in Figure 1, Tone A would have included partials at 1200 and 600 Hz in addition to the higher harmonics). In such a stimulus, a listener who truly perceived the MF as the pitch of Tone B would respond "down," but a listener who perceived the second harmonic would respond "up." Schneider et al. excluded such octave-shifted responses from their analysis altogether, calculating the SI only on the basis of responses that could be clearly classed as F0 or spectral. In keeping with the importance of stimulus variables discussed in the preceding section, Schneider et al. note that octave-shifted responses were given primarily to stimuli with relative high MF values.

[3] There is, unfortunately, a discrepancy between the formula given on page 1242 of Schneider et al.'s paper and the published graphs in the same paper: In the formula, F0 responses are positively poled (i.e.100% F0 responses yields an SI of $+1$), whereas in the graphs, F0 responses are negatively poled (i.e., 100% F0 responses yields an SI of -1). Subsequent work by Schneider and his colleagues (e.g., Schneider and Wengenroth, 2009) has settled on the polarity shown in the graphs, and this is reflected in the formula we use here. Note, though, that this is, in some sense, opposite to the polarity implicit in Seither-Preisler et al.'s AAT. Ultimately, of course, the choice is arbitrary, and for exactly that reason, there is considerable potential for confusion. Caveat lector.

Seither-Preisler et al. took a different approach to inconsistent responses; as noted earlier, they simply excluded many participants whose AAT scores fall in the middle of the range on the grounds that such response patterns cannot be distinguished from guesswork. At the same time, they suggest that such midrange scores might arise for two distinct reasons: Either the participants responded *inconsistently* (that is, giving opposite responses to different presentations of the same stimulus) or they responded *inhomogeneously* (that is, consistently giving F0 responses to some stimuli and spectral responses to others). This is a valuable distinction, especially in light of the clear findings, previously summarized, that certain stimulus variables systematically influence the overall proportion of F0 responses, and in light of Schneider et al.'s finding that some stimulus types seem to yield more octave-shifted percepts. If many individuals exhibit systematic inhomogeneous behavior, then it is obviously an oversimplification to describe everyone as either a spectral listener or an F0 listener.

However, Seither-Preisler et al. were limited in their ability to detect inhomogeneity directly, because their participants heard only a few presentations of each of many stimulus types. Consequently, some of the participants with midrange AAT scores who were excluded for inconsistency might more appropriately have been treated as inhomogeneous. Furthermore, the very notions of inconsistency and inhomogeneity are based on an easily overlooked assumption underlying the MF task itself. Despite its apparent simplicity, the task presupposes that listeners' responses reflect *independent* percepts of the pitch of the two tones in each stimulus. That is, it assumes that listeners perceive the pitch of Tone A and Tone B according to either the MF or the partials, and report a pitch rise or fall across the stimulus on that basis. It does not allow for the possibility that listeners who are asked to report the *direction* of pitch across the stimulus do so on some more holistic basis that does not simply reflect how they perceive static pitch in a single tone (cf. the discussion of contour and interval in Patel, 2008, Chapter 4); as we shall see, there is reason to think that this possibility must be taken seriously. In any case, one of the central goals of the work reported here is a better understanding of response patterns that yield intermediate values of the SI.

Our Studies

Our studies of this topic are ultimately motivated by an interest in individual perceptual and cognitive differences that are potentially relevant to language. However, the focus of the present article is more basic. In order to draw convincing connections between specific individual behavioral differences and other cognitive traits, we will need a well-understood and well-operationalized measure of the behavior in question. As can be seen from the foregoing review, this is precisely what we do not have in the case of the MF task. What we report here is therefore a set of experiments aimed primarily at clarifying what it is that the MF task reveals. Our principal concern is with the distinction between inconsistency and inhomogeneity, and with explanations for midrange SI scores. We also report findings on test–retest reliability, and, in a limited way, we deal with the related issue of the effects of stimulus variables, discussed previously. In keeping with our ultimate interest in individual differences, we also report findings on the influence of three participant variables, namely, age, gender, and musical background.

The data reported in Experiments 1, 2, and 3a come from strictly exploratory experiments. The remaining data, in Experiments 3b, 4a, 4b, and 4c, are drawn from four studies that focused on the relation between the MF task and other perceptual measures relevant to language. The specific issues addressed in the last four experiments have been (or will be) reported elsewhere, and the present article includes only the basic behavioral data from those experiments. Although the studies had different purposes, the same task is used and methodologies are broadly similar. Most importantly, the minor methodological differences in our experiments had no impact on the conclusion that there are two different ways of responding to MF stimuli; indeed, as we have already discussed, experiments in the literature have diverged radically in their methodological choices yet have all converged on this conclusion. It thus made sense to pool the data across the studies, given the benefits of increasing generalizability and statistical power. Detailed discussion of the comparability of the different experiments is provided in the online supplemental appendix.

Method

Stimulus Variables

By and large, our approach to stimulus construction was closer to that of Schneider et al. than to that of Seither-Preisler et al. Within a stimulus, we always held the top frequency of Tone A and Tone B constant, and always kept the harmonic rank of the two sets of partials close (that is, our stimuli resemble the one illustrated in Figure 1). This, in turn, means that the interval between the two missing F0 values was always quite small, between two and four semitones. It also means that the F0 value of the tones was determined entirely by the top frequency and the harmonic rank of the partials, and that the range of F0 values was therefore, especially in our earlier experiments, quite large. In the later experiments (Experiments 4a, 4b, and 4c), influenced by Seither-Preisler et al., we narrowed the range of top frequencies and used lower harmonic ranks, thereby narrowing the range of the missing F0. In the earlier experiments, the tones consisted of three partials, but in the Experiment 4 set, we used a mix of two-partial and three-partial stimuli.

The most significant respect in which our work diverges methodologically from that of Seither-Preisler et al., and especially Schneider et al., is that our experiments involve *fewer stimulus types and more responses to each type*. For example, in Experiment 1, we had only 15 stimulus types, based on a two-dimensional stimulus matrix with five settings of the top frequency and three settings of the harmonic rank of the partials. In each of the 15 cells of this stimulus matrix, every participant gave 10 judgments during the course of the experiment, five in each order of presentation (AB or BA). By comparison, Seither-Preisler et al. had 50 stimulus types, and participants gave only four judgments per stimulus type, two in each order. Schneider et al. had 144 stimulus types and obtained only one response per type; the order of Tone A and Tone B within each stimulus type was randomly assigned. By contrast, all of our data (with minor exceptions due to errors and missing responses and with the systematic exception of Experiment 4b) are based on 10 responses per stimulus type. This gives us a good basis for investigating Seither-Preisler et al.'s distinction between inhomogeneity and inconsistency in participants whose responses are not consistently at one end or the other of the SI scale. Quite unexpectedly, it also allowed us to observe a large order-of-presentation effect in some participants, reported in more detail later, which we believe is relevant to the interpretation of intermediate values of the SI.

Summary of Experiments

Table 1 summarizes the details of all the experiments on which this report is based. All are based on a systematic two-dimensional matrix of stimulus types like the one just exemplified for Experiment 1, with *top frequency* and *spectral composition* (here defined as the harmonic rank of the partials) as the two dimensions of the matrix. The number of stimulus types (i.e., the number of cells in the matrix) varies from 10 to 27. Experiments whose identifiers share a number (e.g., 3a and 3b) have the same stimulus matrix but are not otherwise related. As previously noted, most of the experiments were motivated by research questions beyond the basic goal of clarifying the nature of response patterns in the MF task. These additional questions are summarized in the notes for Table 1.

Participants

Altogether there were 412 participants. In Experiments 1 and 2, participants were mostly friends, colleagues, or family members of one or more of the authors; in Experiment 3a, they were amateur singers from two respected Edinburgh choirs; in Experiment 3b, they were mostly students at Boston University; in Experiment 4a, they were mostly students at the University of Nijmegen; in Experiments 4b and 4c, they were mostly students at the University of Edinburgh. In the experiments involving students, the participants were paid a small sum for participating; in Experiment 3a, a small donation was made to the choirs in which the participants sang. Overall, the great majority of participants were native speakers of English, but there were also native speakers of quite a few other languages as well, in particular, Dutch and Chinese. Except in Experiments 3b and 4c, native language or language background was not a variable we were interested in or one that we attempted to control; except in those two experiments, participants were almost all White Europeans or European Americans, and even in those two experiments, the majority of participants were White native speakers of English or another European language. Participants' ages varied from about 15 to about 75, but in all the experiments that relied on students as participants (3b, 4a, 4b, and 4c), most were in their early 20s. In Experiments 1 and 2, ages were more widely distributed; in the experiment involving choral singers (Experiment 3a), most of the participants were middle-aged or older (median age, 55 years).

Stimulus Preparation

We created our stimuli using an application written for us by Simon Kirby, based on Max/MSP software, which allowed us to specify (a) the F0 and duration of the two tones; (b) the number, harmonic rank, and relative amplitude of the partials; and (c) the duration of the gap between the tones. Except in the second phase of Experiment 4c (see Procedure section), the two tones in each stimulus were 500 ms long, with a 250-ms gap between them, corresponding exactly to Schneider et al.'s stimuli (Seither-Preisler et al. used tones 500 ms long with a 500-ms gap). All our stimulus tones had flat spectra, again following Schneider et al. rather than Seither-Preisler et al.; we experimented informally with modifying the spectral slope and concluded that, except in cases of very steep slope, there was no readily perceptible difference, but we have not done a controlled comparison. Most of the experiments were done with a version of the software that did not control the phase relations of the component partials; Experiment 4c used a newer version in which phase can be controlled, and the stimuli were created with the partials in phase. Note in this connection that Smoorenburg (1970) specifies that the tones in his stimuli were not

Table 1
Summary of Experiments

Experiment	No. of participants	Stimulus matrix: Top frequencies	Stimulus matrix: Spectral composition
1	37 (16 F, 21 M)	300, 500, 900, 1400, 2200 Hz	345/456, 567/678, 689/890
2	20 (12 F, 8 M)	500, 750, 1050, 1400, 1800 Hz	345/456, 678/789
3a[a]	23 (13 F, 10 M)	250, 750, 1050, 1400, 1800, 2000, 2500, 3000, 4000, 5000, 6000 Hz	345/456, 678/789
		300, 500, 900, 1400, 2200 Hz	567/678
3b[b]	152 (105 F, 47 M)	[as 3a]	[as 3a]
4a[c]	50 (39 F, 11 M)	500, 675, 900, 1200, 1600, 2150 Hz	**34/45**, 345/456, **56/67**
4b[d]	73 (57 F, 16 M)	[as 4a]	[as 4a]
4c[e]	57 (41 F, 16 M)	[as 4a]	[as 4a]

Note. Under "stimulus matrix," frequency level is indicated by the top frequency of the stimulus, while spectral composition is indicated by the harmonic rank of all partials in both stimulus tones. The harmonic rank of the partials is specified in abbreviated form as, e.g. 345/456, which is to be read as meaning that Tone A consists of harmonics 3, 4, and 5 of the MF, while Tone B consists of harmonics 4, 5, and 6. In these abbreviated formulas, harmonic 10 (which was used only in Experiment 1) is symbolized as 0. In the Experiment 4 set, two of the spectral composition conditions (in boldface) involve only two partials. MF = missing fundamental.
[a] Participants were amateur choral singers. This experiment was intended to explore findings about musical preferences in Schneider and Wengenroth (2009), but was inconclusive in that respect. [b] This experiment was part of a study on individual differences in a task involving implicit learning of an artificial tone language (Caldwell-Harris, Biller, Ladd, Dediu, & Christiansen, 2012). [c] This experiment was part of a large study on individual differences, with both language-related (e.g., nonword repetition, vocabulary learning) and control tasks (e.g., IQ). Includes test–retest reliability data. [d] This experiment was part of a study on hemispheric differences in pitch processing, to be reported in a separate article. [e] This experiment was part of two separate studies, one on native language and MF perception, and one exploring a possible link between MF perception and the "tritone paradox" (Deutsch, 1991; Repp, 1994).

phase controlled; Seither-Preisler et al. do not specify; P. Schneider (personal communication, August 2012) reports that the stimuli in the Schneider et al. article were phase controlled. Sound files of an illustrative sample of the stimuli can be found in the online supplemental appendix.

Procedure

Most of the experiments were run using an e-prime script written for us by Eddie Dubourg, but Experiments 2 and 4a used a Presentation software (http://www.neurobs.com) script written by Dan Dediu. There were minor variations in the instructions given to participants, but the instructions were always presented in writing on the screen at the beginning of the experiment, and they always involved an explanation that the stimuli consisted of two tones in sequence and that the experimental task was to judge whether the pitch went up or down from the first to the second. The instructions also made clear that this might be difficult to judge and told participants to give their dominant impression. When the stimulus was played, the screen displayed a prompt for a response; participants responded by pressing one of two keys on the keyboard, for example, "U" for "up" and "D" for "down." In Experiment 1, the response trials timed out after 2 s; subsequently, the program was set up so that the participant had to give a response before the next stimulus was presented. All experiments included a short practice session.

Listening conditions varied somewhat: In Experiment 1, some of the participants used a laptop with ordinary headphones or (in a few cases) the laptop's internal speakers, but all of the other experiments were run using professional headphones, either in a booth in a perception lab or using a laptop in a quiet room. Intensity was set at a comfortable level for each listener. In all cases, the stimuli were presented in a blocked random order, fixed between participants in Experiments 2 and 4a, and generated at run time in the others. In all cases, participants heard a randomly ordered full set of stimulus types in both AB and BA order (i.e., two occurrences of each stimulus type in the matrix for a given experiment, one in each order), followed by the same set in a different random order, and so on, until the full set had been presented five times. Participants were given regular opportunities for self-timed breaks; in most of the experiments, these opportunities were offered at the end of each full stimulus set.

Experiment 4b deviated somewhat from the summary just given. As noted in Table 1, this experiment was intended to explore hemispheric differences in pitch processing, and stimuli were presented to one ear with white noise in the other ear. Each stimulus was presented four times in each order to each ear, for a total of 16 times, instead of 10 times, as in the other experiments. We found no link between monaural presentation and the distribution of F0 and spectral responses, and for purposes of the present report, we pool left- and right-ear presentations for each cell in the stimulus matrix.

Experiments 4a and 4c provide two different measures of test–retest reliability. In Experiment 4a, assessing reliability was a specific goal of the larger study: Participants were retested on exactly the same material using exactly the same procedures after an interval of 1 to 2 weeks. As for Experiment 4c, it consisted of three separate blocks of stimuli for two separate studies, all run in a single experimental session lasting approximately 45 min. The first block contained the stimuli for the basic MF task just sketched and served as a baseline for comparison with the other two blocks. The second block contained the stimuli for a separate study on the "tritone paradox" (Deutsch, 1991; Repp, 1994) and is not relevant here, except insofar as it served as a distractor between the first and third blocks. The third block contained a set of stimuli identical in all respects to those of the first block, except that they were much shorter (Tone A and Tone B were each 180 ms long, with a gap between the tones of 20 ms). This manipulation was exploratory, to see if stimulus duration would affect listeners' perception; many participants complained that the short-stimulus task was much more difficult, but in the event, duration had little effect on individual patterns of responses. Consequently, we report the results of the first and third blocks of this experiment as a separate measure of reliability.

Data Reduction and Analysis

The patterns of responses are broadly similar in all seven of our experiments, and, except where specified, the analyses reported here are for all experiments pooled. The retest data from Experiments 4a and 4c are not included in these pooled analyses. A breakdown of the results by experiment is presented in the online appendix. All analyses were conducted using R (R Development Core Team, 2012) and some of its libraries.

We used the SI, introduced in Equation 1, to provide a basic quantitative characterization of each participant's behavior. Given the structure of the stimulus space in our experiments, for each participant we calculated the SI separately for *each cell* in the stimulus matrix and then took the average of the individual cell SIs to arrive at a single overall SI for each participant. As noted previously, the SI ranges from -1.0 to $+1.0$, with 100% F0 responses yielding an SI of -1.0 and 100% spectral responses yielding an SI of $+1.0$. As we pointed out in Footnote 3, this polarity is the opposite of that given in the published formula in Schneider et al.

A substantial minority of participants respond so consistently that there is an SI of either $+1$ or -1 in almost all cells.[4] By definition, participants with overall SI near $+1$ or -1 have a pattern of responses, that is, in the terms suggested by Seither-Preisler et al., both consistent and homogeneous. These are the participants who can be classified confidently as either F0 or spectral listeners. However, because each participant gave 10 responses in each cell of the stimulus matrix, computing the SI for each participant in each cell of the stimulus matrix allowed us to gain a fairly clear idea of the consistency and homogeneity of the responses of participants whose overall SI lies nearer the middle of the scale. We also discovered, unexpectedly, that the intermediate SI values sometimes reflect the presence of an order of presentation effect: Some participants gave different responses depending

[4] In fact, in Experiment 1, we were able to detect a faulty stimulus because, in one cell of the matrix, about a third of the participants had an SI near 0, despite having an SI near $+1$ or -1 in all the other cells. Investigation revealed that Tone A in the AB-order stimulus in that cell had incorrect partials, such that both F0 and spectral listening should yield the same response. This anecdote gives an idea of the consistency and reliability with which some participants respond. On the other hand, the fact that some participants respond very consistently should not blind us to the fact that many others do not.

on whether the two tones of the stimulus were in AB or BA order, and, moreover, the size of this effect differed in different parts of the stimulus matrix.

To better understand the patterning of the cell-level SI and to investigate the structure of the participants' responses, we also carried out a principal components analysis (Jolliffe, 2002) for each experiment separately. We found that the first two principal components are extremely similar across experiments: The first principal component, which explains at least half of the variance, is equivalent to the overall SI, and the second component expresses one of the clear response patterns at intermediate SI values. We converted this second principal component to an index that ranges, like the SI, from -1 to $+1$; we refer to this measure as the consistency index (CI). Further details on our mathematical treatment of the principal components data are given in the online supplemental appendix; further detail on the interpretation of the CI is given in the next section.

Results

Individual Differences Between Participants

Basic distribution of the SI. Our findings replicate those of the studies already discussed, in the sense that we find a range of response patterns from an SI near -1.0 to an SI near $+1.0$. Figure 2 shows the distribution of the SI for all experiments pooled. The distribution does not appear to be Gaussian (Q–Q plot and Shapiro–Wilk normality test, $W = 0.964$, $p < .0001$), but contrary to what is reported, especially by Schneider et al., it is not bimodal either (Hartigan's dip test for unimodality, $D = 0.018$, $p = .48$; Hartigan & Hartigan, 1985). There appears to be a bias toward F0 responses, as suggested by Seither-Preisler et al. There is also a single clear mode about 0, which seems to correspond to the substantial number of participants that Seither-Preisler et al. excluded for "guessing." The difference between our findings and Schneider et al.'s is apparently attributable to the controls for octave-shifted percepts that Schneider et al. included in their study (see previous discussions). P. Schneider (personal communication, August 2012) informs us that if octave-shifted percepts are counted as F0 percepts, then the overall distribution of the SI in their data looks more like what we report here: a greater proportion of F0 listeners and a more Gaussian shape.

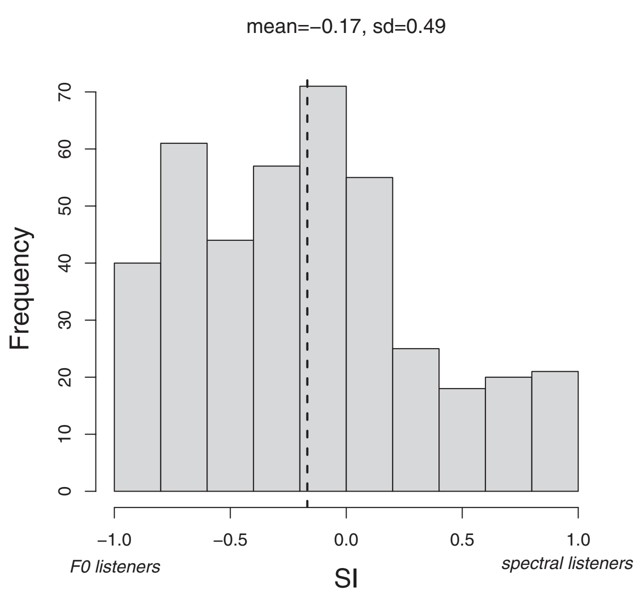

Figure 2. Distribution of Schneider index (SI; -1.0 = F0 listener, $+1.0$ = spectral listener) for all 412 participants in the seven experiments (retest data from Experiments 4a and 4c excluded); dashed vertical line represents the mean. This figure may be compared with Figure 1b in Schneider et al. (2005).

Principal components analysis. Principal components analysis, carried out on each experiment separately, confirms the validity of the overall SI as the principal expression of the individual differences on the MF task. In every experiment, the first principal component, explaining between 46.5% and 80.3% (mean 65.7%) of the between-participants variance, is effectively equivalent to the SI. As noted earlier, a common second principal component also emerged for all experiments, explaining between 6.5% and 18.4% (mean 12.1%) of the variance, with eigenvalues ranging between 0.47 and 3.24 (mean 1.05), and correlating only weakly with the first component ($r = -.10$, $p < .05$). Principal components other than the first two were inconsistent across experiments, had eigenvalues less than 0.5 except in Experiments 3a and 3b, and were not obviously interpretable. Further detail on the principal components analysis is given in the online supplemental appendix.

Interpretation of consistency index. Inspection of the data showed that the second principal component reflects a consistent difference between responses in the lower and higher portions of the top frequency dimension of the stimulus matrix. Closer investigation of the data from participants with SI values in the middle of the scale showed that some give predominantly spectral responses at lower frequency levels and predominantly F0 responses at higher levels. Figure 3, which shows the responses of one such participant, illustrates this pattern graphically. This specific pattern of responses seems to be what is captured by the CI. Importantly, the CI is skewed toward positive values (minimum, -0.53; maximum, 0.96; median, 0.09; mean, 0.13). Only two participants of the entire group of 412 had CI values as low as approximately -0.5; removing them increased the minimum to -0.28. This suggests that one theoretically possible pattern of responses, namely, F0 responses to stimuli with lower overall frequency and spectral responses to those with higher overall frequency, occurs only very rarely, whereas the opposite pattern is quite common.

Order of presentation effect. In addition to the pattern expressed by the CI, we found, to our surprise, that a number of participants show an order of presentation effect that interacts with frequency level. Specifically, at low frequency levels, AB stimuli (where the MF falls) are more likely than BA stimuli (where the MF rises) to elicit F0 responses, whereas at high frequency levels, the reverse is true. An alternative way of stating this observation, which may ultimately provide more insight into its cause, is to say that low frequency level favors "down" responses and high frequency level favors "up" responses. Inspection revealed that this tendency does not affect all participants equally: Some show no influence of order of presentation at all, whereas others show dramatic differences between low and high frequency level. Overall, however, the pooled data reflect this tendency, as can be seen in Figure 4; as can also be seen, frequency level affects the response to BA stimuli more than the response to AB stimuli.

For the purpose of further analyses, we quantified this order effect by taking the mean of the absolute value of the difference in the SI between AB and BA responses at each frequency level. There are weak but significant correlations between the order effect, so quantified, and both the SI and CI—for the SI, $r = .19$, $p < .001$, Spearman's $\rho = .30$, $p < .001$; for the CI, $r = -.20$, $p < .001$.

Test–retest reliability. As explained in the Procedure section, we report two different assessments of the test–retest reliability, not only for the SI but also for the CI and the order effect. In Experiment 4a (a conventional test of reliability involving exact repetition of the experiment after an interval of 1 to 2 weeks), the correlation between test and retest for the SI was $r = .87$, $p < .001$.[5] The CI also showed high test–retest reliability ($r = .83$, $p < .001$); the order effect slightly less so ($r = .52$, $p < .001$). In Experiment 4c, the second test—presented later in the same experimental session, as described in the Procedure section—involved stimuli that differed in duration but were otherwise identical to those of the first test. Here the correlation between test and retest was $r = .94$, $p < .001$, for the SI; $r = .73$, $p < .001$, for the CI; and $r = .74$, $p < .001$, for the order effect. It therefore seems clear that the individual differences tapped by the MF task are very robust.

Effect of Stimulus Variables

Overall frequency level and spectral composition. We saw in the introduction that both Schneider et al. and Seither-Preisler et al. report more F0 responses to stimuli with overall higher frequency, though it is not easy to tell whether the effect is primarily due to higher frequency, as such, or to higher harmonic rank of the partials (spectral composition, in our terms). Our experiments strongly suggest that actual frequency is the more important factor, though the order effect illustrated in Figure 4 may be relevant as well. The left-hand panel of Figure 5 shows the effect of frequency level on the overall distribution of responses for all experiments pooled. (Frequency levels above 2200, which were used only in Experiments 3a and 3b, and which yielded an overwhelming preponderance of F0 responses, are excluded.) It can be seen that at very low frequency levels (500 Hz and below, implying MF values between 30 and 100 Hz), the distribution of the SI is nearly Gaussian, with a mode near 0, that is, showing no preference for either spectral or F0 responses. The distributions then become flatter, with considerable numbers of clear F0 listeners and spectral listeners in the midrange of frequencies, around 1000 Hz. As the frequency level increases from the midrange, the distribution becomes increasingly skewed toward F0 responses.

For comparison, the right-hand panel of Figure 5 presents a similar analysis, showing the effect of spectral composition on patterns of responses. (It is based only on the Experiment 4 set, which shared the same spectral composition variables.) It is difficult to interpret as clearly as the left-hand panel, and there is certainly no obvious trend. As discussed in the introduction of this article, there is a significant degree of interdependence between the stimulus variables, and our data do not permit us to explore this issue further.

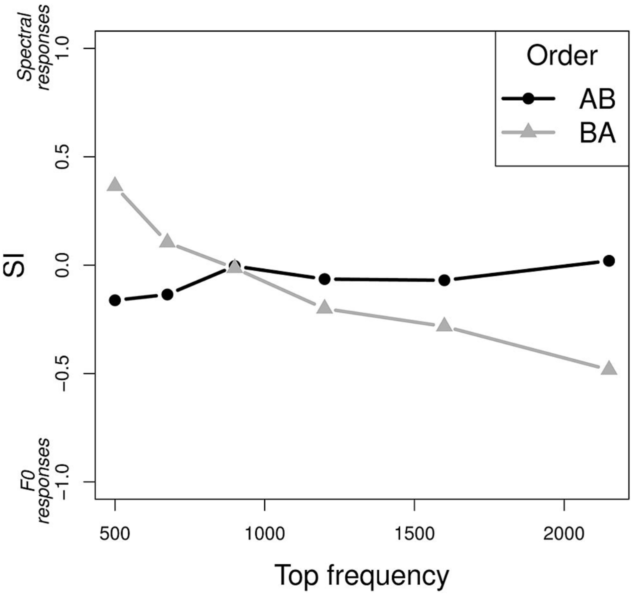

Figure 4. Interaction between order of presentation and frequency level, as shown by pooled data from the Experiment 4 set. Data from the other experiments (which have different specific frequency levels) are qualitatively very similar. It can be clearly seen that BA stimuli show a tendency to elicit more F0 responses at higher frequency levels, whereas AB stimuli do not. This interaction is highly significant (ANOVA, $F(5, 2148) = 3.3 \times 10^{29}$, $p < .001$).

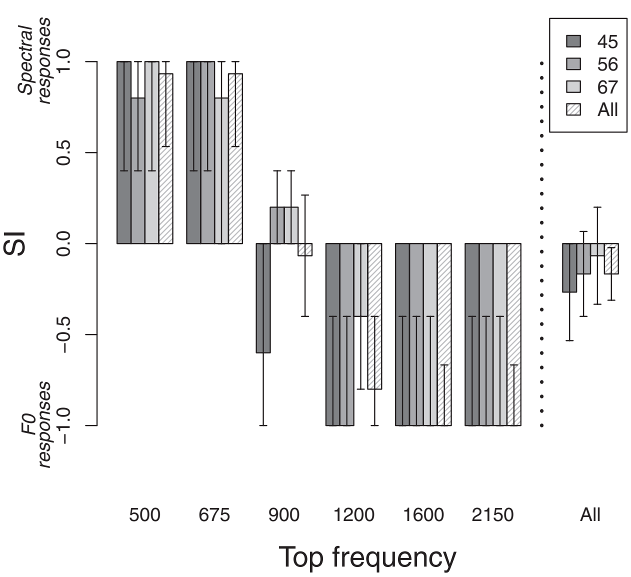

Figure 3. Individual responses of one "inhomogeneous" participant from Experiment 4c, representative of roughly 7.5% of participants who give mostly spectral responses to stimuli with low frequency level and mostly F0 responses to those with high frequency level.

[5] This may be compared with Schneider et al.'s (2005, p. 1046) report of test–retest reliability of $r = .96$ for a subgroup of 37 participants retested 6 months after the original experiment.

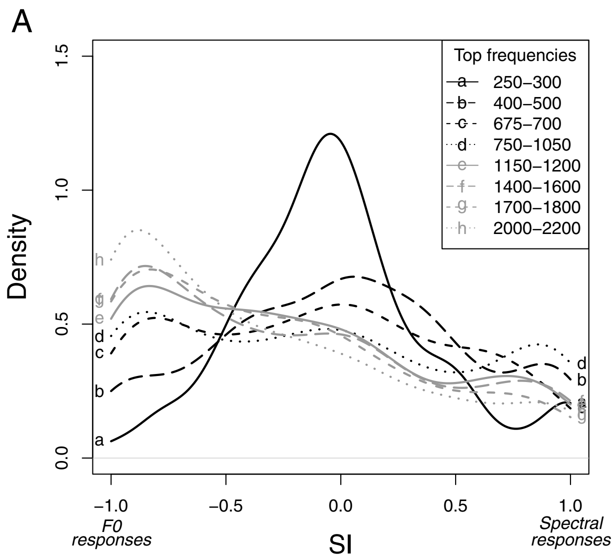

Figure 5. Effects of overall frequency level and spectral composition on patterns of responses. The lines represent envelope-like approximations of the distributions, computed by kernel density estimation (Silverman, 1986). Panel A shows clearly that higher frequency levels give rise to more F0 responses; this panel may be compared with Figure 1c in Schneider et al. (2005). Panel B is more difficult to interpret. See text for further discussion.

Cluster Analysis

Given the apparent diversity of response patterns, we subjected the data to a *k*-means cluster analysis, locating every participant in a three-dimensional space defined by SI, CI, and the order effect. We found that seven clusters fit across all experiments. (Technical details are given in the online supplemental appendix). Figure 6 shows these clusters plotted in two different two-dimensional

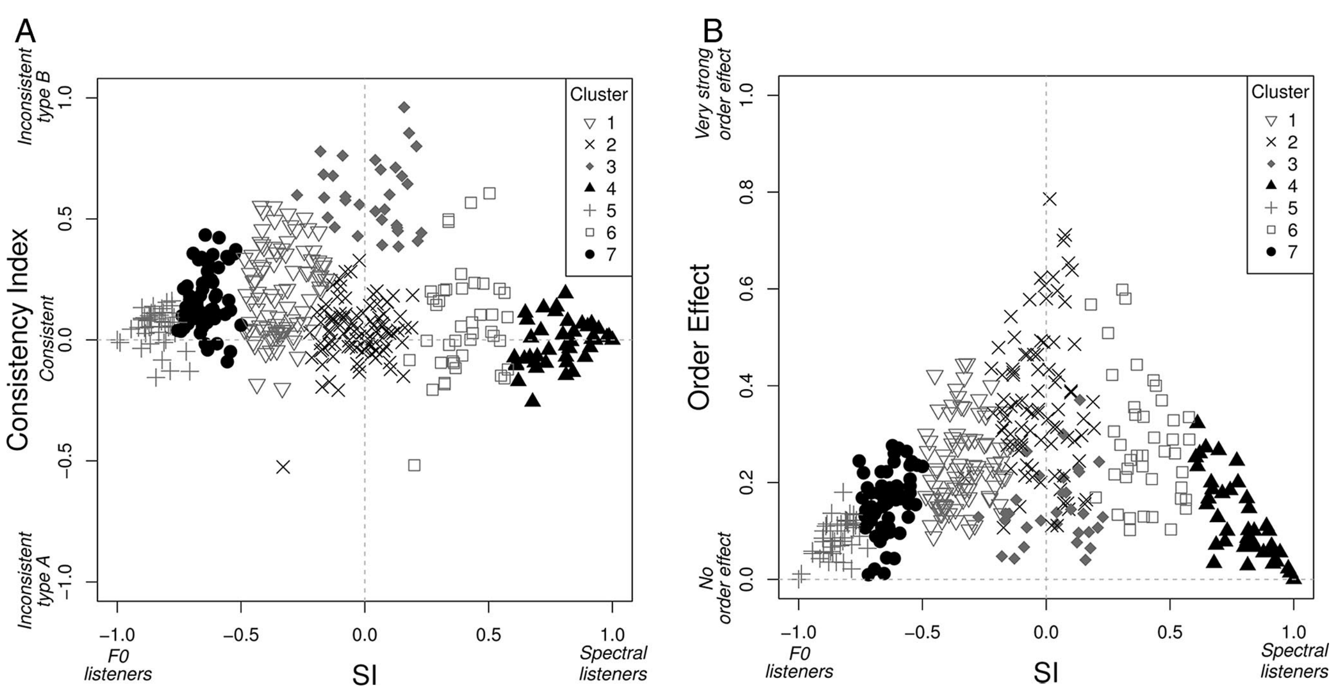

Figure 6. Scatterplots of the 7 *k*-means clusters for our data. Only 2 two-dimensional projections are shown here: Panel A plots the Schneider index (SI) against the consistency index, and Panel B shows the SI plotted against the order effect. See text for further discussion.

projections, one showing the relation between the SI and CI, and one showing the relation between the SI and the order effect. The clusters are clearly interpretable. Cluster 4 (black triangles) comprises consistent spectral listeners and Cluster 5 (gray crosses) comprises consistent F0 listeners. Cluster 6 (gray rectangles) comprises weak spectral listeners, and Clusters 7 (black circles) and 1 (gray inverted triangles) comprise weak F0 listeners, some of whom show clear effects of frequency level on their pattern of responses. Cluster 2 (black *x*s) comprises listeners with no clear preference and/or those strongly affected by the order effect; Cluster 3 (gray diamonds) comprises listeners who show a strong shift of listening preference from lower to higher frequency levels.

Note that roughly a quarter (22.8%) of all participants fall into Cluster 2 (listeners with no clear preference), about the same proportion excluded as inconsistent by Seither-Preisler et al. However, it can be seen from Figure 6 that this cluster is the one most strongly affected by the order effect. That is, for many of these participants, the fact that the SI is near zero results from a consistent pattern of responses—but one which cannot be expressed by the SI and which, as we suggested earlier, may actually undermine the assumptions underlying the MF task. Further research is clearly called for.

Effect of Participant Variables

Recall that both Schneider et al. and Seither-Preisler et al. were interested in the effects of musical training. Unlike either of those studies, we had very few professional musicians among our participants, but we did collect self-reports on musical activity and training, and, on this basis, we can very roughly classify our participants as either musical (generally corresponding to Seither-Preisler et al.'s amateur musicians) or not (Seither-Preisler et al.'s nonmusicians). Like Schneider et al., we find no effect of musicality on the SI, $t(388.48) = -1.27$, $p = .20$, but there are significant effects on the order effect, $t(424.12) = 5.99$, $p < .001$, with musical participants being significantly less susceptible to order effects than nonmusical participants. Musical participants were also slightly less influenced by frequency level, as expressed by the CI, $t(361.48) = -4.49$, $p < .001$. This is at least consistent with the idea that musical listeners are performing the task as intended, that is, hearing Tone A and Tone B separately and judging their relative pitch level, whereas nonmusical listeners may be treating the pair of tones as some sort of holistic unit.

Schneider and Wengenroth (2009) found no effect of age or gender on the SI. We find no effect of gender, but a slight effect of age, with older participants slightly more likely to give spectral responses. On the full set of 412 participants, the correlation between age and the SI is $r = .16$, $p < .01$; to compensate for the very skewed age distribution in our participant group, we ran the same analysis based only on participants over 25 years of age and found the same tendency but with too little statistical power to reach significance ($r = .20$, $.05 < p < .10$). One can imagine a variety of explanations for an age effect on the SI; probably the most plausible is one based on physical changes in the inner ear, but cortical changes cannot, by any means, be ruled out (cf. Whitfield, 1980). We find no effect of age or gender on either the CI or the susceptibility to the order effect.

Discussion and Conclusion

Our investigations have confirmed that there are robust individual differences in the perception of MF stimuli. As we pointed out in the introduction, a comparison of two recent studies (Schneider et al., 2005, and Seither-Preisler et al., 2007) suggests that these differences will emerge even from experiments that substantially diverge methodologically. In the present set of studies, we have specifically demonstrated that individuals' responses are unaffected by large differences in stimulus duration, and our test–retest reliability results confirm that individuals' responses are consistent over time. At the same time, we have shown that it is something of an oversimplification to classify individuals as "F0 listeners" or "spectral listeners."

First, we have shown that as many as a quarter of all individuals appear to have no consistent response preference at all. Superficially, this finding diverges from the results of Schneider et al. (who report a strongly bimodal "U"-shaped distribution that appears to justify a binary classification of participants), but as noted earlier, the difference may be explainable by Schneider et al.'s careful control of octave-shifted percepts. Our findings more obviously agree with those of Seither-Preisler et al., who excluded roughly a quarter of their participants from further analysis on the grounds that they were guessing. In separate analyses associated with Experiment 4c, we found that some of these inconsistent listeners also respond inconsistently to stimuli used to investigate the "tritone paradox" (Deutsch, 1991; Repp, 1994), which might suggest that their responses reflect a more general difficulty with judging pitch or pitch direction, or alternatively, perhaps, a susceptibility to octave-shifted percepts. In any case, the conclusion that such individuals are merely "guessing" seems decidedly premature because, as illustrated in Figure 6, many of them actually do show a consistent response pattern that simply happens not to be captured by data reduction in terms of the SI. This group of listeners needs to be treated separately in drawing conclusions about MF perception and may be interesting to study in its own right.

Second, we have confirmed and extended others' findings that certain stimulus variables have predictable effects on responses to MF stimuli. The effect of overall frequency level is strong enough that 7.5% of individuals (in our analysis, those in Cluster 3) give consistently opposite responses in different areas of the stimulus space, responding as "spectral listeners" at low overall frequencies and as "F0 listeners" at high overall frequencies. Studying such listeners may provide useful insight into the sources of the two different modes of perceiving MF stimuli. Even among participants who are not so strikingly affected, we have shown that, in general, responses are influenced by the overall frequency level and perhaps by the spectral composition of the stimuli. This has implications for the construction of appropriate stimuli in further research.

Our findings on stimulus and participant variables in MF perception should make it possible for researchers whose interest is in the physical and psychophysical foundations of the phenomenon to make more confident methodological choices, and may help shed light on apparent discrepancies in the results of different studies. It should now also be possible to use the MF task with greater methodological confidence in studies that are not essentially concerned with the phenomenon itself, but with what it tells us about individual differences more generally. For example, it could be revealing to determine the heritability of modes of MF perception, or to investigate the relationship between MF perception and other perceptual and cognitive tasks, from basic auditory sensitivity to language-related tasks such as nonword repetition and digit span. It may also be interesting to investigate brain structure and function (as in the studies by Patel & Balaban, 2001, and Schneider et al., 2005) with a more fine-grained characterization of individual behavioral differences than simply "F0 listener" and "spectral listener." We believe we have provided the research community with a better-calibrated tool for all these purposes.

References

Bernstein, J. G., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *Journal of the Acoustical Society of America, 120,* 3916–3928. doi:10.1121/1.2372451

Caldwell-Harris, C., Biller, A., Ladd, D. R., Dediu, D., & Christiansen, M. (2012, March). *Musical ability and prior tone language experience facilitate learning an artificial tone language*. Paper presented at the American Association for Applied Linguistics, Boston, MA.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception, 8,* 335–347. doi:10.2307/40285517

Faulkner, A. (1985). Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discrimination? *Journal of the Acoustical Society of America, 78,* 1993–2004. doi:10.1121/1.392656

Gockel, H. E., Carlyon, R. P., & Plack, C. J. (2010). Combining information across frequency regions in fundamental frequency discrimination. *Journal of the Acoustical Society of America, 127,* 2466–2478. doi:10.1121/1.3327811

Gockel, H. E., Plack, C. J., & Carlyon, R. P. (2005). Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch. *Journal of the Acoustical Society of America, 118,* 3783–3793. doi:10.1121/1.2126823

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of Statistics, 13,* 70–84. doi:10.1214/aos/1176346577

Houtsma, A. J. M., & Fleuren, J. F. M. (1991). Analytic and synthetic pitch of two-tone complexes. *Journal of the Acoustical Society of America, 90,* 1674–1676.

Jolliffe, I. (2002). *Principal component analysis*. New York, NY: Springer Verlag.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, 7,* 128–134. doi:10.1007/BF02156143

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America, 89,* 2866–2882. doi:10.1121/1.400725

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Bingley, UK: Emerald Group.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America, 77,* 1853–1860. doi:10.1121/1.391936

Moore, B. C. J., & Gockel, H. E. (2011). Resolvability of components in complex tones and implications for theories of pitch perception. *Hearing Research, 276,* 88–97. doi:10.1016/j.heares.2011.01.003

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D., & Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nature Neuroscience, 4,* 839–844. doi:10.1038/90557

Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America, 41,* 1526–1533. doi:10.1121/1.1910515

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Repp, B. H. (1994). The tritone paradox and the pitch range of the speaking voice: A dubious connection. *Music Perception, 12,* 227–255. doi:10.2307/40285653

Ritsma, R. J. (1962). Existence region of the tonal residue. I. *Journal of the Acoustical Society of America, 34,* 1224–1229. doi:10.1121/1.1918307

Ritsma, R. J. (1963a). Existence region of the tonal residue. II. *Journal of the Acoustical Society of America, 35,* 1214–1245. doi:10.1121/1.1918679

Ritsma, R. J. (1963b). On pitch discrimination of residue tones. *International Audiology, 2,* 34–37. doi:10.3109/05384916309070127

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., . . . Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience, 8,* 1241–1247. doi:10.1038/nn1530

Schneider, P., & Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review, 28,* 315–328. doi:10.1080/07494460903404402

Schouten, J. F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Akademie van Wetenschap, 41,* 1086–1093.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 743–751. doi:10.1037/0096-1523.33.3.743

Silverman, B. W. (1986). *Density estimation*. London, UK: Chapman and Hall.

Smoorenburg, G. F. (1970). Pitch perception of two-frequency stimuli. *Journal of the Acoustical Society of America, 48,* 924–942. doi:10.1121/1.1912232

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55,* 1061–1069. doi:10.1121/1.1914648

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research, 1,* 155–182. doi:10.1016/0378-5955(79)90025-X

Whitfield, I. C. (1980). Auditory cortex and the pitch of complex tones. *Journal of the Acoustical Society of America, 67,* 644–647. doi:10.1121/1.383889

Received June 25, 2012
Revision received September 12, 2012
Accepted November 2, 2012 ■