Comparing $F0$ discrimination in sequential and simultaneous conditions (L)

Christophe Micheyl[a] and Andrew J. Oxenham
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(Received 1 December 2004; revised 12 April 2005; accepted 18 April 2005)

In an influential study, Carlyon and Shackleton [J. Acoust. Soc. Am. **95**, 3541–3554 (1994)] measured listeners' performance (d') in fundamental-frequency ($F0$) discrimination between harmonic complex tones (HCTs) presented simultaneously in different spectral regions and compared their performance with that found in a sequential-comparison task. In this Letter, it is suggested that Carlyon and Shackleton's analysis of the simultaneous-comparison data did not adequately reflect their assumption that listeners were effectively comparing $F0$'s across regions. A reanalysis consistent with this assumption is described. The new results suggest that under the assumption that listeners were effectively comparing $F0$ across regions, their performance in this task was substantially higher than originally estimated by Carlyon and Shackleton, and in some conditions much higher than expected from the performances measured in a traditional $F0$-discrimination task with sequential HCTs. Possible explanations for this outcome, as well as alternative interpretations, are proposed.
[DOI: 10.1121/1.1929228]

PACS number(s): 43.66.Ba, 43.66.Fe, 43.66.Hg [RAL] Pages: 41–44

I. INTRODUCTION

Signal detection theory (SDT) provides a unifying framework for comparing results across studies involving different psychophysical paradigms, tasks, or procedures (Green and Swets, 1966; Jesteadt and Sims, 1974; Jesteadt and Bilger, 1975). Such comparisons may then be used to gain more insight into the mechanisms of perception than could be obtained with comparisons restricted to data all gathered using the same paradigm.

An interesting illustration of this approach can be found in an influential study by Carlyon and Shackleton (1994). In that study, the same listeners were tested in two experiments that were both thought to involve fundamental-frequency ($F0$) discrimination abilities. One experiment (experiment 3a in Carlyon and Shackleton's article) was a basic two-interval two-alternative forced choice (2I2AFC) $F0$-discrimination experiment, in which listeners were asked to indicate which of two consecutive harmonic complex tones (HCTs) had a higher $F0$. The other experiment (experiment 3b in Carlyon and Shackleton's article) involved comparing two consecutive pairs of simultaneous HCTs filtered into different spectral regions. In one of the two pairs, the two HCTs had the same $F0$; in the other pair, the two HCTs had different $F0$'s. The listener's task was to indicate in which of the two pairs the two simultaneous HCTs had different $F0$'s.

In order to compare performance in the two experiments, Carlyon and Shackleton transformed the proportion of correct responses into d'. They found that, when the two HCTs being compared both contained peripherally resolvable harmonics, the d''s measured in the experiment with simultaneous HCTs were not significantly different from those predicted based on the results of the experiment with only sequential HCTs. In contrast, when one of the two simultaneous HCTs contained resolved harmonics and the other did not, the d''s measured in the experiment with simultaneous HCTs were significantly lower than predicted. Carlyon and Shackleton noted that this pattern of results was consistent with the hypothesis that the $F0$'s of resolved and unresolved harmonics are processed by different mechanisms, the outputs of which cannot be directly compared. They suggested that the necessary "translation" of these outputs into a common format caused performance in $F0$ comparisons between resolved and unresolved harmonics to be limited by some internal "translation noise." Because of its important implications for pitch perception theories, Carlyon and Shackleton's study has generated much interest in recent years (Meddis and O'Mard, 1998; Grimault *et al.*, 2002; Plack and Carlyon, 1995; Gockel *et al.*, 2004; Micheyl and Oxenham, 2004).

In this note, we suggest that Carlyon and Shackleton's SDT analysis of the results of their experiment with simultaneous HCTs is not consistent with their assumption that listeners were comparing $F0$'s across regions in that experiment. We propose a reanalysis of the experimental data that is consistent with this assumption and, on the basis of the results of this reanalysis, point out alternative interpretations of the data.

II. SIGNAL-DETECTION-THEORETIC CONSIDERATIONS

Central to the present reanalysis is the question of what strategy listeners could use for optimal task performance in Carlyon and Shackleton's two experiments.[1] The optimal

[a]Electronic mail: cmicheyl@mit.edu

(likelihood-ratio) strategy determines the relationship between d' and proportion correct in the considered experiment.[2]

For the sequential $F0$-discrimination experiment, the answer is obvious. Since this experiment involved a basic two-interval, two-alternative forced-choice (2I2AFC) paradigm with a roving standard, the optimal strategy is to subtract the perceived $F0$ in the first interval from that in the second interval, and respond "interval 2" if the resulting difference was larger than zero or "interval 1" otherwise (Green and Swets, 1966; Macmillan and Creelman, 1991). With this decision rule, the relationship between d' and the proportion of correct responses, P_C, is given by Green and Swets (1966) (see also: Macmillan and Creelman 1991)

$$d' = \sqrt{2}\Phi^{-1}(P_C), \tag{1}$$

where Φ^{-1} denotes the inverse of the cumulative standard normal.

The optimal strategy in the experiment involving two pairs of HCTs is less obvious. Although this experiment superficially involved a 2I2AFC task, each observation interval in fact contained two (simultaneous) HCTs. In one of the observation intervals, the two $F0$'s were the same, while in the other they were different. If it is assumed that the listeners have access to the $F0$'s of the two HCTs in each interval, and that they compare these $F0$'s, the experiment is more accurately described as a dual-pair comparison or four-interval AX (4IAX) task.

Possible strategies for the 4IAX paradigm have been described in earlier publications (Macmillan *et al.*, 1977; Kaplan *et al.*, 1978; Noreen, 1981; Rousseau and Ennis, 2001). The strategies available to listeners depend on the specifics of the experimental design. If the design is such that it prevents listeners from relying on absolute judgments of the observations, and instead forces them to rely on comparisons between observations, then the best that listeners can do is compare the absolute value of the differences between the two observations in each pair, and select the pair for which the absolute difference is larger; this is the so-called differencing strategy for the 4IAX paradigm (Rousseau and Ennis, 2001).

Two implementation features of Carlyon and Shackleton's experiment with simultaneous HCTs constrained which strategy the listeners could use in order to perform optimally. First, the baseline stimulus $F0$ was roved widely across trials, which drastically limited the use of a fixed internal reference, and produced highly correlated sensory observations. Second, in the stimulus pair containing HCTs with different $F0$'s, the higher-$F0$ HCT was assigned randomly to the lower or to the higher spectral region; thus, the sign of the difference between the two observations in each pair was irrelevant to the task. It can be demonstrated that under these circumstances, the differencing strategy is the optimal strategy (Micheyl and Messing, unpublished).

Under the differencing strategy, the relation between proportion correct and d' in the 4IAX paradigm is (Macmillan *et al.*, 1977)

$$P_C = [\Phi(d'/2)]^2 + [1 - \Phi(d'/2)]^2, \tag{2}$$

where Φ is the cumulative standard normal. Different (but equivalent) formulations of the relationship between proportion correct and d' for the differencing strategy in the 4IAX design can be found elsewhere (e.g., Rousseau and Ennis, 2001).

Thus, the relationship between d' and proportion correct in Carlyon and Shackleton's experiment using simultaneous pairs of HCTs should, in theory, be that described in Eq. (2). Instead, Carlyon and Shackleton treated both experiments as standard 2I2AFC paradigms and, accordingly, used Eq. (1) or an equivalent of it, in order to transform into d' the proportion-correct values measured in their two experiments. However, the relationship between d' and proportion correct described by Eq. (1) is based on the assumption that listeners made only two observations and only one comparison on each trial, which is inconsistent with Carlyon and Shackleton's tacit assumption that listeners compared $F0$ across spectral regions in each of the two observation intervals from each trial. Accordingly, we reanalyzed the data from Carlyon and Shackleton's experiment involving two pairs of harmonic complexes using Eq. (2).

III. REANALYSIS OF CARLYON AND SHACKLETON'S DATA

The d' values shown in Fig. 7 of Carlyon and Shackleton's article were converted to proportions of correct responses using the inverse form of Eq. (1). The resulting proportions were then transformed back into d' using the inverse form of Eq. (2).[3] The recomputed d' values are plotted in Fig. 1 as open symbols connected by solid lines. These can be compared to the d' values originally calculated by Carlyon and Shackleton for this task, which are shown here as filled symbols connected by solid lines. The d' values that were predicted by Carlyon and Shackleton based on the results of the sequential $F0$-discrimination task are also replotted here as open symbols connected by dotted lines.

It can be seen that the recomputed d' values are systematically and markedly ($\sim 80\%$) larger than the d' values originally calculated by Carlyon and Shackleton. This suggests that if listeners really compared $F0$ across spectral regions in the simultaneous case, Carlyon and Shackleton's (1994) calculations substantially underestimated performance in this task. Furthermore, when the recomputed d' values are compared to those predicted based on the results of the basic sequential $F0$-discrimination experiment, it is found that in three of the four conditions tested, the predictions fall below the recomputed d' values. Leaving aside the MIDHI 88-Hz condition (for which Carlyon and Shackleton indicated a strong reason to expect higher performance, based on pitch-pulse asynchrony cues), planned comparisons (two-way repeated-measures ANOVAs contrasting the predicted and recomputed d' values, with the $F0$ difference included as a factor) showed a significant difference between the predicted and the recomputed d' values for the LOMID 250-Hz condition [$F(1,2) = 921.054, p = 0.01$], as well as for

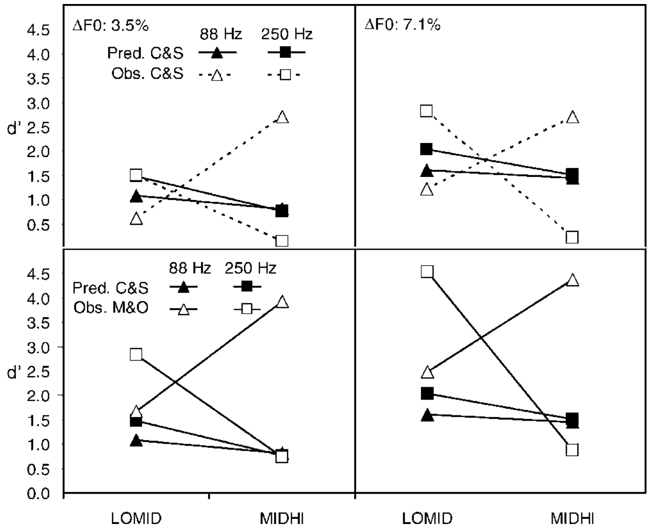

FIG. 1. Comparison between the d' values calculated by Carlyon and Shackleton (1994) and those calculated in the present reanalysis. The top panels show the observed and predicted d' values replotted from Carlyon and Shackleton (1994). The format of these panels is the same as that of their Fig. 9: the "observed" d' values, which were calculated based on the measured proportion of correct responses in the experiment with simultaneous HCTs, are shown as filled symbols connected with solid lines; the "predicted" d' values, which were calculated based on the results of the sequential $F0$-comparison experiment, are represented by empty symbols connected with short-dashed lines. Triangles are used to represent data obtained with a nominal $F0$ of 88 Hz, squares to represent data obtained with a nominal $F0$ of 250 Hz. The left-hand panel shows data obtained using an $F0$ separation ($\Delta F0$) of 3.5%, while the right-hand panel shows data obtained with a $\Delta F0$ of 7.1%. The spectral regions into which the stimuli were filtered are indicated in the abscissa. The lower panels show the recalculated "observed" d' values that resulted from the present reanalysis of Carlyon and Shackleton's data. The recalculated values are shown as empty symbols connected by solid lines. Carlyon and Shackleton's original predictions are also replotted here to facilitate comparisons; they are the same as in the upper panels, and are again shown as filled symbols connected by solid lines.

the LOMID 88-Hz condition [$F(1,2)=94.881, p=0.010$], but not for the MIDHI 250-Hz condition [$F(1,2)=0.108, p=0.774$].

IV. DISCUSSION

The present results can be interpreted in two main ways, depending on whether or not one is willing to accept that Carlyon and Shackelton's experiment with two pairs of simultaneous HCTs filtered into different spectral regions involved basically the same $F0$-comparison mechanisms as the more traditional sequential $F0$-discrimination experiment. If it is assumed that this was the case, then the results of the present reanalysis indicate that simultaneous $F0$ comparisons between HCTs filtered into different spectral regions can be significantly more accurate than $F0$ comparisons between sequential HCTs filtered into the same spectral region, even when one of the two simultaneous complexes contains resolved harmonics while the other does not (see the results of the LOMID 88-Hz condition). This conclusion is the opposite of that drawn by Carlyon and Shackleton (1994), who concluded (based on the same data) that performance was systematically *poorer* than predicted in conditions involving simultaneous comparisons between resolved and unresolved harmonics.

Nevertheless, the new analysis remains consistent with Carlyon and Shackleton's main finding that $F0$ comparisons between resolved and unresolved harmonics are less accurate than comparisons between harmonics of the same resolvability status. Indeed, even with the recomputed data shown in Fig. 1, the amount by which observed performance exceeds predicted performance is somewhat larger in resolved–unresolved conditions (LOMID-88 Hz and MIDHI-250 Hz) than in resolved–resolved conditions (LOMID-250 Hz). This is consistent with relatively more noise being present in the former than in the latter conditions, regardless of whether the sequential listening conditions involved more noise than the simultaneous ones or not (see footnote 3 in Carlyon and Shackleton). Two possible reasons for which simultaneous $F0$ comparisons may be more accurate than sequential comparisons are, first, that sequential comparisons involve memory noise while simultaneous comparisons do not (Durlach and Braida, 1969; see footnote 3 in Carlyon and Shackleton, 1994), and second, that instantaneous across-region comparisons are immune to correlated noise across peripheral channels (Durlach *et al.*, 1986; Dai and Green, 1992). Another possible explanation, which is more specific to the considered experiments by Carlyon and Shackleton, is that in these experiments the $F0$ of the stimuli varied over time (it was modulated over 5% at a rate of 2.5 Hz). This variation in $F0$ over time may have limited listeners' ability to make sequential $F0$ comparisons, as it may have required that the estimated $F0$ tracks from the two HCTs be temporally aligned prior to being compared. In contrast, when the HCTs to be compared were simultaneous, the $F0$ modulation being phase coherent, listeners could compare $F0$'s across spectral region on a moment-by-moment basis. This could explain the higher performance in simultaneous across-region comparisons than in sequential within-region comparisons.

While there are some reasons to expect sequential $F0$ comparisons between HCTs filtered into the same spectral region to involve sources of internal noise to which simultaneous $F0$ comparisons between HCTs filtered into different regions may be immune, there are as many if not more reasons to expect the converse. For instance, the salient timbre differences between complexes filtered into different spectral regions may have a detrimental effect on the ability to make fine $F0$ discrimination between these complexes (e.g., Faulkner, 1985; Moore and Glasberg, 1990; Micheyl and Oxenham, 2004). Furthermore, due to pitch-shift effects, differences in spectral region may cause complexes having the exact same $F0$ to sound like they have a different pitch (Chuang and Wang, 1978; Singh and Hirsh, 1992). Finally, it has recently been shown (Gockel *et al.*, 2004) that the perception of the $F0$ of a harmonic complex filtered in a spectral region can be adversely affected by the simultaneous presence of another complex in a different region, especially when the $F0$'s of the two complexes are relatively close. In the light of these earlier results, the conclusion that performance in simultaneous $F0$ comparisons across different spectral regions is generally better than performance in sequential $F0$ comparisons within the same region may appear difficult to accept.

An alternative interpretation of the current results is that the two considered experiments of Carlyon and Shackleton involved different perceptual mechanisms. Although both experiments probably required the extraction of $F0$-related information at some stage, it is quite possible that this $F0$ information was utilized very differently in the two experiments. For instance, task performance in the experiment with simultaneous complexes could be based primarily on the output of $F0$-based grouping/segregation processes, whereby simultaneous HCTs having different $F0$'s were perceived as less well-fused than HCTs having the same $F0$ (Darwin, 1992). Although grouping and segregation processes possibly involve some implicit comparison of $F0$ information across spectral regions, they may not necessarily engage the same $F0$-comparison mechanisms as a sequential task.[4] Like Carlyon and Shackleton's (1994) original analysis, the present reanalysis is based on the assumption that the sequential and simultaneous "$F0$-discrimination" experiments involved the same type of $F0$-comparison mechanisms. Without this assumption, it becomes unclear how performance obtained in these two types of tasks can be meaningfully compared in order to elucidate the nature of the underlying pitch mechanisms.

In any case, the present reanalysis and the resulting reversal of the conclusions regarding relative task difficulty illustrate the importance of selecting the correct decision models when comparing performance across different tasks.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 05216). The authors are grateful to Robert Carlyon, Laurent Demany, Joshua Bernstein, David Messing, and an anonymous reviewer for helpful comments on earlier versions of the manuscript.

[1]These two experiments are described as experiment 3a and experiment 3b in Carlyon and Shackleton's (1994) article.

[2]Throughout this article, d' is meant as the standard index of performance from signal-detection-theory, traditionally defined in reference to the basic yes/no (YN) paradigm as the distance beteen the means of the two (Gaussian) internal probability density functions corresponding to the two possible stimulus alternatives. This d' is sometimes referred to in the psychophysical literature as the "true" d'. One important feature of d', when it is defined in this way is that, all other things being equal, its value will not depend on the experimental paradigm that is used to measure it.

[3]The inverse form of Eq. (2) can be found by completing the square. When this is done, it is found that $d' = 2\Phi^{-1}(1/2 + \sqrt{x})$, with $x = P_C/2 - 1/4$. P_C, d', and Φ^{-1} are as defined above. Note that the domain over which this formula is applicable is $0.5 < P_C < 1$.

[4]In fact, there is some evidence that the auditory grouping/segregation mechanisms related to the perception of mistuning operate partly independently from pitch-perception mechanisms. For instance, the grouping of simultaneous frequency components appears to be governed by spectral regularity, not just harmonicity (Roberts and Brunstrom, 2003). Other arguments for the notion that the detection of mistuning between simultaneous tones may involve other processes than explicit pitch comparisons can be found in Demany and Semal (1992).

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms," J. Acoust. Soc. Am. **95**, 3541–3554.

Chuang, C. K., and Wang, W. S. (**1978**). "Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order," J. Acoust. Soc. Am. **64**, 1004–1014.

Dai H., and Green, D. M. (**1992**). "Auditory intensity perception: Successive versus simultaneous, across-channel discriminations," J. Acoust. Soc. Am. **91**, 2845–2854.

Darwin, C. J. (**1992**). "Listening to two things at once," in *The Auditory Processing of Speech, from Sounds to Words*, edited by M. E. H. Schouten (Mouton de Gruyer, New York), pp. 133–145.

Demany, L., and Semal, C. (**1992**). "Detection of inharmonicity in dichotic pure-tone dyads," Hear. Res. **61**, 161–166.

Durlach, N. I, and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63–72.

Faulkner, A. (**1985**). "Pitch discrimination of harmonic complex signals: Residue pitch of multiple component discrimination?" J. Acoust. Soc. Am. **78**, 1993–2004.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (**2002**). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," Percept. Psychophys. **64**, 189–197.

Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.

Jesteadt, W., and Sims, S. L. (**1975**). "Decision processes in frequency discrimination," J. Acoust. Soc. Am. **57**, 1161–1168.

Kaplan, H. L., Macmillan, N. A., and Creelman, C. D. (**1978**). "Tables of d' for variable-standard discrimination paradigms," Behav. Res. Methods Instrum. **10**, 796–813.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, UK).

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "The psychophysics of categorical perception," Psychol. Rev. **84**, 452–471.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Micheyl, C., and Messing, P., "Likelihood ratio, optimal decision rules, and correct-response probabilities in the 4IAX paradigm".

Micheyl, C., and Oxenham, A. J. (**2004**). "Sequential $F0$ comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms," J. Acoust. Soc. Am. **116**, 3038–3050.

Moore, B. C. J., and Glasberg, B. R. (**1990**). "Frequency discrimination of complex tone with overlapping and non-overlapping harmonics," J. Acoust. Soc. Am. **87**, 2163–2177.

Noreen, D. L. (**1981**). "Optimal decision rules for some common psychophysical paradigms, in *Mathematical Psychology and Psychophysiology (Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics, Vol. 13)*, edited by S. Grossberg (American Mathematical Society, Providence, RI), pp. 237–279.

Plack, C. J., and Carlyon, R. P. (**1995**), "Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics" J. Acoust. Soc. Am. **98**, 1355–1364.

Roberts, B., and Brunstrom, J. M. (**2003**). "Spectral pattern, harmonic relations, and the perceptual grouping of low-numbered components," J. Acoust. Soc. Am. **114**, 2118–2134.

Rousseau, B., and Ennis, D. M. (**2001**). "A Thurstonian model for the dual pair (4IAX) discrimination method," Percept. Psychophys. **63**, 1083–1090.

Singh, P. G., and Hirsh, I. J. (**1992**). "Influence of spectral locus and $F0$ changes on the pitch and timbre of complex tones," J. Acoust. Soc. Am. **92**, 2650–2661.