Disruptive Effects of Delayed Auditory Feedback on Motor Sequencing

by

Steven Adams Finney

B.A., Computer and Information Science,

University of California, Santa Cruz, 1980

B.A., Linguistics, University of California, Santa Cruz, 1990

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in the Department of Cognitive and Linguistic Sciences

at Brown University

May 1999

Abstract of "Disruptive Effects of Delayed Auditory Feedback on Motor Sequencing" by Steven Adams Finney, Ph.D., Brown University, May, 1999.

Lashley (1951) argued that accounting for the correct execution of long sequences of planned actions is one of the central problems of cognition. One major question is the role of feedback (i.e., perceptual consquences of action) in the execution of such sequences. Although Lashley argued that rapid sequences must be centrally programmed, the robust impairment caused by delayed auditory feedback (DAF) has been used to argue for a role of feedback in tasks such as speech, musical performance, and tapping. There is, however, a bit of a paradox. Although DAF robustly impairs performance in all these domains, auditory feedback itself appears not to be necessary for an adequate level of performance (although fine control may suffer). In fact, the impairing effects of delayed auditory feedback have remained a puzzle, inadequately explained by standard theories of stimulus chaining or feedback guidance. In this dissertation, I report on a set of experiments involving delayed auditory feedback and rhythmic tapping, and show that the results are incompatible with two more recent proposals for DAF, the Displaced Rhythm Hypothesis of Howell, Powell, and Khan (1983), and the Node Structure Theory of MacKay (1987). I then advance a new class of DAF theories based on the idea that subjects are attempting to entrain to the delayed stimuli.

This thesis by Steven A. Finney is accepted in its present form by
the Department of Cognitive and Linguistic Sciences
as satisfying the dissertation requirement for
the degree of Doctor of Philosophy

Date ....
William H. Warren, Jr., Director

Recommended to the Graduate Council

Date ....
Sheila Blumstein, Reader

Date ....
Bruce Kay, Reader

Date ....
Bruno Repp, Reader

Approved by the Graduate Council

Date ....
Peder J. Estrup
Dean of the Graduate School and Research

Vita

Steven A. Finney

EDUCATION

Ph.D., Cognitive and Linguistic Sciences, Brown University, May, 1999.

B.A., Linguistics, University of California, Santa Cruz, May, 1990

B.A., Computer and Information Science, University of California, Santa Cruz, May, 1980

ACADEMIC HONORS

University Fellowship, Brown University, 1992-1993.

College Honors, Crown College, UC Santa Cruz, 1990.

Phi Beta Kappa, UC Santa Cruz, 1990.

ACADEMIC AND WORK EXPERIENCE

1998: Research Assistant, Michael Tarr. Maintained and enhanced RSVP experimental software package for the Macintosh.

1993-1997: Teaching Assistant, Brown University. Classes included Introduction to Statistics, Introduction to Cognitive Science, Visual Perception, Semantics, Human Cognition, and Cognitive Psychology Laboratory.

1980-1989: UNIX Systems Programmer, The Santa Cruz Operation, Santa Cruz, CA.

PUBLICATIONS

Finney, S.A. (1997). Auditory feedback and musical keyboard performance. *Music Perception*, 15, 153-174.

Finney, S.A., Protopapas, A., and Eimas, P.D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language*, 35, 893-909.

Protopapas, A., Finney, S. A., and Eimas, P. D. (in press). Baseline conditions in structural induction: Comments on Pitt, Smith, and Klein (1998). Journal of Experimental Psychology: Human Perception and Performance.

Protopapas, A., Finney, S.A., and Eimas, P.D. Attentional allocation to syllabic boundaries in English first-syllable stress words. Submitted for publication.

Protopapas, A., Finney, S.A., and Eimas, P.D. (1995). Effects of syllabic position in the perception of spoken English. Proceedings of Eurospeech '95, Madrid.

PRESENTATIONS

"Delayed Auditory Feedback and Rhythmic Tapping", 9th annual meeting of the New England Sequencing and Timing Association. Mar 6, 1997.

"Auditory feedback and musical keyboard performance", 7th annual meeting of the New England Sequencing and Timing Association. Feb 8, 1997.

"Auditory feedback and musical keyboard performance", invited colloquium at Rhode Island College. April 9, 1997.

"Investigating syllabic effects in the on-line processing of English". Poster by Allopenna, P.D., Protopapas, A., Finney, S.A., and Eimas, P.D.. CUNY Sentence Processing Conference. May 21-23, 1996.

RESEARCH INTERESTS

Music performance, auditory feedback, sequential motor (finger) behavior, sentential context effects on lexical access, on-line semantic integration in sentence processing.

Acknowledgments

I've been here for a long time, so I get to have a lot of acknowledgements...

For this dissertation, I chose a 4 person committee, and not only that, I knew in advance that each one of them was actually going to read and criticize what I gave them. More than one former graduate student told me that this was a dangerous situation to put myself in, but, as I had expected, the end result was immeasurably improved by all their contributions...although it was quite an experience going through 4 sets of detailed comments page by page!

I'd like to thank Bill Warren for jumping in as thesis advisor for a somewhat random and weird project, just because it could be called "motor control" (what behavior can you observe that *doesn't* involve motor control?). He's provided valuable insight throughout, and all references to "entrainment" or "oscillators" are due to him.

I'd like to thank Sheila Blumstein for being my advisor on an earlier planned psycholinguistics dissertation, and then sticking around as a committee member on my current dissertation. Thinking of her reading over my should has caused me to try to be very precise about my references to speech. And when she pointed out that my writing was unclear, it's usually because my thinking was unclear.

I thank Bruce Kay for encouragement at very early stages when I was doing some exploratory music studies (which later developed into an article in Music Perception), and for providing insights and statistical advice throughout.

I thank Bruno Repp for being helpful and encouraging throughout, and for being a role model as a bit of a maverick researcher (changing from a highly productive career in speech research to doing music research because it was what he wanted to do), and for being able to come up with a reference on almost anything, although too often in French or German. He also has a wonderful sense of elegant writing style in English.

A number of faculty were not on my committee, but have been very helpful, and I'd like to acknowledge them. Thanks to Mike Tarr for lab space and encouragement, Bill Heindel for access to the psychology subject pool, Steven Sloman for pressing me on the issues addressed in Chapter 7, and Jim Anderson for...well, for being Jim Anderson! Thanks also to Russ Church, for encouragement and unstinting advice. I think the questions I'm dealing with are fairly different from his, but it's been invaluable to make use of his knowledge and experience. Polly Jacobson was wonderful to work with on earlier projects. Although I ended up not working on language for my dissertation, I think good linguistics is a wonderful approach to argumentation, and she's one of the best.

Last, but not least, I'd like to thank Peter Eimas. The work in this dissertation is a direct descendent of a paper I wrote in the proseminar my first semester at Brown, entitled "Motor Programming, Serial Behavior, and Wind Instrument Fingering". Peter didn't think much of it at the time, but apparently he found me redeemable, since I'm still here. Peter trained me as a careful experimentalist; we've been coauthors on a few papers, for which I feel honored.

And now I'd like to go back to Bill Warren. Protocol required that I mention him first, and so I did. But it's hard to describe how much it's been a genuine pleasure working with him: it's been useful, educational, challenging, and fun for me, and I hope it was the same for him.

Outside of Brown, both Caroline Palmer and David Rosenbaum have been very helpful, providing both comments and encouragement.

The graduate students during my time at Brown have been great to be around. I'd like to expressly mention a few of them. David Ascher encouraged me to learn Python; I dragged my feet on it for a year, but he finally talked me into it. Many of the analyses in this dissertation probably couldn't have been done without it. Thanassi Protopapas provided friendship and lots of arguments. Gary Strangman and Andrew Duchon, my former long-time office mates, both made comments about me in their dissertation acknowledgments which were simultaneously complimentary and insulting (reflecting our general office vibe). Not that I went out of my way to look at what they said about me or anything. I, however, am above all that, and would merely like to sincerely thank them for their friendship and assistance, and to say that I look forward to Gary's finally explaining how it *really* all works together, and to Andrew constructing a robot which can do something interesting (perhaps starring in a screenplay entitled "Ho-bot!"). I'd also like to remind them of the saying, "He who finishes last finishes best". Thanks also to more recent officemates Jeff Saunders for the peace and quiet, and Aparna Nadig for the gummis.

Many thanks to Bob Fifer for keeping the computers running!

On the other coast, I'd like to thank Donna Karolchik and Dan Auvil for their friendship throughout the years, and give a remembrance to Amy Snader, who might have beaten me to a Cognitive Science PhD if she'd had the chance. Finally, thanks to the the Sons of Champlin and the Malik family of Ochrid, Macedonia, for providing a significant part of the soundtrack to this dissertation.

Contents

List of Figures

Chapter 1

Introduction

Most observable aspects of human behavior are in the form of motor acts, often complex, precisely timed *sequences* of motor acts. Speech and musical performance are two interesting cognitive behaviors that can be viewed from this general perspective of "serial behavior". Lashley (1951) viewed such sequencing as one of the central problems of cognition, and argued that the data support a central planning view (what is now often called "motor programming") rather than a stimulus-response chain explanation. Keele (1987) has suggested that the human capability to flexibly organize complex sequences of behavior may be the primary capability distinguishing humans from other animals, with language being one often-cited instance of this flexibility. Thus, the investigation of serial behavior is critical to the study of human cognition.

One very general question is, how are such sequences, whether sentences, sonatas, or swimming, actually produced? More specifically, how are transitions made from one unit to the next? How are such sequences timed properly? How are planned sequences executed correctly, in the face of both noise internal to the system, and changes in the external environment? One approach to addressing these issues focusses on the role of "feedback": the manner in which the perceived consequences of action are used to direct further action. In early work, feedback played a role in two types of models. One such model was the stimulus-response chain approach, in which the feedback from one act serves as the stimulus for the next act in the sequence. In this view, feedback is a necessary part of any sequence execution. A second approach viewed feedback as having a role in error correction, with feedback providing an error signal that is used to engage in corrective behavior in a "closed-loop" system. As a third possibility, feedback might have no important role in behavior at all, with sequences of action being completely under central control.

Two experimental approaches have been used to address the role of feedback. One is to investigate the effects of *removal* of the auditory feedback information; if behavior changes, that would indicate a dependence on feedback in the normal course of events (and detailed investigation of the impairment may indicate what aspects of behavior are guided by the feedback). A second approach is to investigate the consequences of *altering* feedback (that is, what the impact is on behavior when the perceived feedback information is "incorrect" in some fashion).

This dissertation is concerned with one particular channel of feedback, audition, and with tasks in which such feedback is likely to serve an important role. The two cognitive behaviors mentioned above, speech and music, are examples of such tasks, since in both instances the primary output of the system is sound. Within the auditory feedback literature, one interesting (and much investigated) form of altered feedback is *delayed* auditory feedback (DAF), in which the auditory feedback to a subject is delayed by a small amount, e.g., 250 ms. Two early studies (Lee, 1950; Black, 1951) reported independently that speech under DAF conditions was severely impaired, showing a significant slowing of rate, an increase in sound level, and an increase in articulatory errors (most often insertions or repetitions). A large number of studies since then (see Yates, 1963, for an early review and references) have replicated this effect, and the impairment is fairly striking to anyone who has observed or experienced it.

Two concerns motivated the early DAF work. One was the general interest at this time in cybernetics and feedback control theory (e.g., Lee, 1950a; Lee, 1951; and Fairbanks, 1954, all cite Wiener's *Cybernetics*); the impairment found under DAF was assumed to indicate the breakdown of a normal feedback guidance system for speech. The second motivation was clinical, with two different aspects. First, since the impairment under DAF had similarities to stuttering, it was thought that DAF might provide clues to the source of stuttering. Second, it was thought that the robust impairment demonstrated under DAF might serve as a test for hearing ability in cases where subjects with no physical hearing impairment nonetheless demonstrated behavior characteristic of hearing loss ("auditory malingering" or "psychogenic deafness"). Because DAF causes impairment which is hard to avoid, even with practice and effort, it might serve as a clinical test for such circumstances (see, e.g., Tiffany and Hanley, 1952).

This issue of clinical testing seems to have motivated the earliest specific studies of finger tapping and DAF, although impairment of simple rhythmic tasks under DAF was noted in the earliest studies, e.g., Lee (1950b). Classification of errors in speech is time consuming, and requires trained evaluators. Tapping of rhythmic patterns (for example, on a Morse code key) can be more easily evaluated in a clinical situation. Several studies (e.g., Chase et al., 1959, 1961b, 1961c) showed that tapping with DAF was impaired in a way qualitatively similar to the disruption of speech (i.e., slowed rate, increased key force, and insertion of taps), even at relatively low sound levels of DAF.

Musical performance also shows impairment under DAF. Havlicek (1968) reported that both errors and loudness increased under DAF in performance on a variety of musical instruments. Gates and Bradshaw (1974) found increases in an elapsed time measure with DAF. Finney (1997) reported that the predominant error under DAF was insertion of notes, again similar to the reports for speech and tapping.

Thus, DAF reliably causes impairment in a variety of serial tasks, with qualitatively similar impairment (e.g., insertions, increased loudness, and slowed rate) in each area. Insertions are particularly striking: they are not merely a change in degree, but are an actual additional action not part of the original plan. Not only are these insertions conceptually noteworthy, they are also the predominant error. Such insertions will be the major focus of this dissertation.

Strikingly, in spite of hundreds of studies of DAF, no adequate account of DAF impairment has emerged. Garber and Siegel (1982, p 96), wrote that "no satisfactory account has yet been offered as to why DAF should be disruptive to any speaker". Zimmerman et al. (1988, p 438) wrote "In spite of considerable research effort and in spite of its powerful effects on speech, no single (or satisfactory) explanation of DAF has emerged". MacKay (1987, p 178) wrote "the last fifteen years have seen virtually no studies of DAF, [although] not because we now understand the phenomenon...". DAF is therefore interesting in and of itself as a robust and unexplained phenomenon. However, understanding its effects may also shed light on the role of feedback guidance, and perhaps even on foundational issues of how action and perception are bound together.

This dissertation is an investigation of the effects of DAF in the domain of finger tapping. I have chosen to investigate tapping for a number of reasons. It is related to musical performance (my primary area of interest), in that it involves *finger* movements producing sound; a particularly close analogy to tapping is keyboard playing. Single finger tapping is, however, a much simpler task than the complex keyboard performance investigated in Finney (1997); it involves only simple up and down movement of a single finger. This simplicity makes it easier to include precise experimental controls than with speech or music, and the resulting data are also easier to evaluate. Finally, current computer methodology allows relative ease of data collection, stimulus control, and analysis, allowing collection of a larger body of data and more precise analysis than was possible in earlier DAF tapping studies. Since, as noted, DAF effects on tapping are qualitatively quite similar to the effects on speech and music, tapping may provide the easiest way to look at the DAF phenomenon. This idea is not new: over 35 years ago, Yates (1963b, p 111) suggested that "it may at this stage be advantageous to concentrate on activities which are simpler than speech, but which are disrupted by DAF, such as rhythmic tapping".

In current experimental psychology, investigating an empirical phenomenon such as DAF is typically done within the context of a particular theory, where some aspect of the phenomenon is felt to provide a test of some aspect of that theory. Although I explicitly evaluate a number of particular theories later, I don't believe that any of them has much claim to plausibility in this domain. Thus, on some level I view this dissertation as simply addressing (with an open mind) the question "What causes DAF impairment?". However, a number of interesting theoretical issues guide the investigation, including the following:

1. Central vs Peripheral Control: The most relevant theoretical issue involving DAF in the literature is that of central or peripheral control: whether action is totally preprogrammed, or whether it is dependent on or guided by peripheral feedback signals. The literature has often viewed this as an all-or-none other dichotomy, although current thinking seems to be that components of both are involved (Delcomyn, 1980). For instance, feedback is almost certainly necessary for a system to respond to environmental change. It also appears to be necessary for some levels of fine control. What, exactly, does feedback control, and under what circumstance? How is this control implemented?

2. Feedback Guidance vs Interference: The DAF effect has often been assumed to provide data about feedback guidance systems. However, a foundational issue is whether the DAF effect relates to feedback guidance at all. That is, the impairment caused by DAF might be due to a feedback guidance or error correction system which is being misled by false information, or it could be a result of some other factor, e.g., "interference" due to a particular pattern of timing overlap between executed actions and a stimulus sequence. Addressing this question in detail requires, on the one hand, providing a detailed theory of how such a feedback system would work, and on the other hand explaining the precise process of interference, along with an explanation of why DAF in particular causes such interference.

3. Is DAF impairment in speech and tapping in fact the same phenomenon? There are a huge number of differences between these tasks (with respect to complexity, cognitive meaning, continuity or discreteness, sources of information, richness of the units of sequencing). Does the same explanation apply to each? Will further investigation of impairment details in tapping reveal further similarities or differences?

1.1 Thesis Organization

The organization of this dissertation is as follows. Chapter 2 selectively reviews the existing empirical results on both auditory feedback absence and delayed auditory feedback in speech, tapping, and music. Chapter 3 reviews and critiques existing theories that have been proposed to account for the phenomenon. Three particular theories are chosen for further investigation, and interesting and crucial predictions made by these theories are provided. Chapter 4 then describes a preliminary tapping experiment, which confirms that the present technique replicates previous results.

Chapter 5 tests one crucial question relating to the delay of maximal impairment (the "critical interval"). Chapter 6 investigates precise timing details which are relevant to two specific theories. Chapter 7 addresses the question (raised above) of feedback guidance vs interference, and Chapter 8 raises the possibility of a new class of theory based on entrainment. Chapter 9 summarizes the results, evaluates the questions raised in the introduction, and provides a number of questions for future research. Two Appendices provide some additional details; they will be referenced occasionally throughout the dissertation.

Chapter 2

Empirical Review

This chapter provides an empirical review of the existing studies on auditory feedback, describing data from studies involving adults engaged in speech, musical performance, and tapping. Feedback conditions include normal feedback, feedback absence and feedback alteration (primarily DAF), as well as extraneous stimuli. These findings will then serve as the background for the later theoretical discussion.

Some issues which inform this review are the following:

1. Is auditory feedback necessary for performance? If so, what impairment occurs when it is absent?

2. What are the symptoms of DAF impairment? Are they similar across different types of activities, or do they differ? If similar, it would suggest that a common process is involved in all cases.

3. Perhaps DAF impairment occurs because subjects are not hearing what they "expect". If so, how does DAF impairment differ from hearing extraneous auditory stimuli?

2.1 Speech

The effects of delayed auditory feedback (DAF) on speech were the object of extensive investigation in the 1950's and 1960's, beginning with the early reports of Lee (1950a, b) and Black (1951). A basic DAF speech experiment involves having a subject speak into a microphone (e.g., reading a written passage aloud), while hearing their own speech amplified over headphones and delayed by some amount (e.g., 200 ms).[1] The primary reported symptoms are a decrease in subjects' speech rate (slowing down), articulatory errors (predominantly repeated speech sounds, sometimes described as resembling stuttering), an increase in speech loudness, an increase in voicing time, and an increase in fundamental frequency (see Yates, 1963, for a review of the early literature).

One aspect of speech and speech feedback should be noted. Controlled auditory feedback manipulations usually involve alteration of the air-conducted feedback reaching the ears; this is what is manipulated when the speech from a microphone is delayed and played back through headphones. However, speech also involves a significant component of bone-conducted auditory feedback, which is *not* affected by this manipulation. Thus, in a standard DAF speech experiment, there is necessarily a synchronous or immediate auditory feedback component in addition to the DAF. I will use the term SAF in this dissertation for such "Synchronous Auditory Feedback".

---

[1] Throughout this review section, I provide all delay times in milliseconds to facilitate comparisons, even when the original source specifies the data in some other unit. In some cases, the figures provided here may imply a precision which is not in the original studies (e.g., 200 ms instead of .2 seconds.)

2.1.1 Feedback Absence

Given the presence of bone conducted feedback, implementing complete auditory feedback absence with speech is non-trivial (perhaps impossible). However, there are two situations which come close. One involves studies using high levels of masking noise (which is assumed to mask both air-conducted and bone-conducted feedback). The other arises in cases of adventitious deafness, in which adult fluent language speakers lose their hearing. In either case, the term auditory feedback "absence" may be taken as an oversimplification of the situation.

Masking

One approach to the study of feedback absence has been to use very high levels of masking noise (sometimes described as being near the pain threshold), in hopes of blocking the bone-conducted feedback. It is well known that masking nose increases the *loudness* of speech (the Lombard effect, for review, see Lane and Tranel, 1971). Masking seems to cause little or no increase on articulatory errors. The effects of masking noise on speech rate are somewhat mixed (see Lane and Tranel, 1971, and Siegel et al., 1982, for two differing claims). Klein and Wolitzky (1970) used 100 dB of masking noise, and found that various measures of social/communicative function were impaired in free expressive speech; they do not provide explicit details on articulatory impairments or rate. The key result for present purposes is that there is no reported evidence of articulatory errors in this condition.

Adventitious Deafness

Occasionally, fluent adult speakers will lose their hearing due to some form of disease or trauma. Lane and Tranel (1971) provide references suggesting that "fluent articulation [is maintained] after experimental or traumatic deafening". Borden (1979) describes speakers who become deafened after acquiring language as suffering "little effect on speech", except possibly after a long period of deafness. Cowie and Douglas-Cowie (1983) report that adventitiously deafened subjects have high intelligibility, particularly those speakers who became deaf after the age of 18. However, they do report somewhat slowed speech (partly a result of vowel lengthening). Although they report some deletions of prefixes and suffixes, insertions are very rare. Waldstein (1990) found that speakers who became deaf in adulthood were lacking in "phonetic precision" with respect to normals, suggesting that "auditory feedback [is used] to calibrate fine phonetic adjustments". However, such speakers had high intelligibility and maintained phonological distinctions, and there appears to be no indication of the gross impairment found under DAF.[2]

Both Waldstein (1990) and Leder et al (1987) provide data showing that such subjects speak slower than normals. This slowed speech rate is of some interest, since it also corresponds with one of the reported DAF symptoms.

Overall, although these speakers have a slowed rate of speech, there seems to be no indication of any articulatory impairment corresponding to that found under DAF. In addition, even such subtle impairments as do occur may only develop over time (Waldstein's subjects had been deaf for 7 to 40 years), and thus may exaggerate impairment due to feedback absence itself. Although the study of Leder et al. didn't find any differences between those subjects who had been deaf for less than 5 years, and those who had been deaf for a longer period, this still leaves a long time for development of effects. Thus, auditory feedback does not appear to be *necessary* for basic spoken performance.

---

[2]In using this description, I am basically assuming that the subphonetic errors referenced in, e.g., Waldstein, are a different class of errors than insertions, i.e., that feedback absence does not cause a certain "gross" kind of error.

2.1.2 Delay Intervals and Impairment

Switching now to studies of DAF, one frequently reported finding is that the delay interval which causes maximal impairment (in terms of duration and error measures) in adults is about 200 ms; this "critical interval" (or "delay of maximal impairment"; I will freely use both terms) is important to a number of the theories reviewed later. Since this critical interval will be particularly important in later discussion, I will cover these studies in some detail. What is particularly interesting is that delays *greater* than this critical interval cause less impairment.

Black (1951), in a study testing 11 evenly-spaced delay values between 0 and 300 ms found that the duration effect (increased time to speak a fixed passage) was maximal at 180 ms, following an inverted U-shaped curve (later studies, e.g., Fairbanks, 1955, have described this curve as positively skewed). Fairbanks (1955) and Fairbanks and Guttman (1958) report a study using delays of 0, 100, 200, 400, and 800 ms seconds. Correct word index (number of correct words per second), articulation errors, and duration impairments were maximal at 200 ms. Fairbanks (1955) also observes that of the articulation errors, the *percentage* of addition errors (insertions and repetitions) was maximal at 200 ms and states that "the most distinctive characteristic of peak disturbance is the high incidence of additions". Butler and Galloway (1957) tested 0, 85, 170, 255, and 340 ms delays and found maximal error impairment at 170 ms. MacKay (1968) used delay intervals of 0, 100, 200, 263, 375, and 750 ms in a sentence repetition task, and reported that correct syllable interval (a form of rate measure) and articulatory errors (repetitions) were maximal at 200 ms. Howell, Powell, and Khan (1983) found maximal impairment at 150 or 200 ms in an elapsed time measure for experiments in which subjects read 40 item lists whose members were isolated vowels. Howell and Archer (1984), however, found peaks (with an elapsed time measure) closer to 300 ms for higher levels of amplification. Fabro and Daro (1995) tested monolingual Italian subjects at delays of 0, 150, 200, and 250 ms; 150 ms produced the slowest rate, while 200 ms produced the highest number of errors.

Spilka (1954a), however, did not find significant differences in impairment for a range of delays between 94 and 375 ms, using syllable duration, phonation time, and loudness as measures; this seems to be the only study *not* finding such an effect. It is perhaps noteworthy that errors were not one of the variables measured.

Except for this last study the finding that there is a delay (or range of delays) around 200 ms which is maximally impairing for speech (with less impairment for delay values greater than *or* less than that delay) seems highly robust. As first noted by Black (1951), and commented on by many others, it is noteworthy that this delay interval approximately corresponds to values for syllable length (Black, 1951 takes 220 ms to be the mean duration of a syllable; other authors cite Lenneberg (1967)'s report of 6 syllables per second in connected speech (i.e., 166 ms/syllable), while Huggins (1968) reports 180 ms as the *median* duration of a syllable in read text). This correspondence has sometimes been taken to support the idea that the syllable is the unit of motor control for speech. It should be noted that syllable length is necessarily highly variable, depending not only on speech rate but on the exact consonants and vowels involved; any theory based on syllable length will need to take this into account.

Interestingly, two other common measures of DAF impairment do not peak at 200 ms, but rather reach an asymptote at 100 ms, and stay at the same level for higher delays. These two measures are intensity (loudness) and fundamental frequency (Fairbanks, 1955). Intensity (corresponding roughly to key force for tapping), although important for some theories, is not robust in the current experiments and will not be a major topic. Fundamental frequency will not be discussed further as it doesn't have any obvious tapping analogy, and because it may be confounded with intensity in any case.

The critical interval is interesting for a number of reasons. Many theories, as will be seen later, predict that impairment should continue to increase with increasing delays, but this doesn't occur.

2.1.3 Delay values and tempo

Given the peak of maximal impairment reported above, one question is whether the delay of maximal impairment is affected by speech *rate* (as might be expected, e.g., in a theory in which correction occurred done on-line on the basis of some form of discrete unit not specifically defined by duration, e.g., the syllable, word, or phoneme). A second question (which can be confused with the first) is whether speech rate affects the *amount* of impairment. Comparing the amount of impairment for a given delay at different speech rates does not of itself directly address the question of the delay which causes maximal impairment. This confounding second question will be addressed first.

Impairment and speech rate

Kodman (1961), using an error measure, found less impairment with 200 ms delay at slow reading rates than at faster reading rates. In particular, no errors occurred when the reading rate was slowed to 72 words/minute. MacKay (1969) reported a higher probability per syllable of stuttering (repetition) at a fixed delay of about 200 ms when subjects spoke at a maximal rate than when they spoke at their normal rate. MacKay (1970), also using a "probability of stuttering" measure, found that subjects stuttered less under DAF when speaking at their normal rate than at a maximum rate, and MacKay (1968) (see below) also found less impairment at slower rates. Overall, the results strongly suggest that DAF impairment is lessened at slow speech rates, which can be taken as an instance of speed-accuracy tradeoff.

Maximal impairment interval and speech rate

In contrast to the above studies showing an effect of speech rate on amount of DAF impairment, a number of studies have suggested that the delay of maximal impairment is constant over speech rate. MacKay (1968) had adult subjects repeat sentences at three rates ranging from .32 sec/syllable to .86 sec/syllable, and with two different delays (200 and 263 ms); the suggested manner of slowing down was "prolonging speech sounds without pausing". At all three rates, fewer articulatory errors (in this case, repetitions) occurred with the 263 ms delay than with the 200 ms delay. There was less overall impairment at the slower rates.

Huggins (1968), using duration measures and a fine phonetic analysis in a single subject found that the delay interval of maximal impairment (tested across a wide range, with 10 ms increments) did not noticeably vary between a sentence composed only of short syllables (less than 180 ms) and a sentence composed only of long syllables (greater than 300 ms); in both cases, the critical interval for these measures was near 250 ms. Repetitions didn't occur at all below 150 ms delay, and were maximal at about 250 ms.

Siegel, Fehst, Garber, and Pick (1980) tested adults with a sentence repetition task at 0, 250, 375, 500, and 625 ms delay; subjects spoke both at a normal rate and as fast as possible. Siegel et al. reported that the interval of maximal impairment for duration measures was 250 ms for both rates (articulatory errors were minimal). There were no significant effects of speech rate (thus, in this case, no speed-accuracy tradeoff). Finally, is worth noting that the interval of maximal impairment changes with age, being longer in children than in adults (MacKay, 1968; see also Siegel, 1980).

Although the above studies report no shift in the delay of maximal impairment with speech rate, Robinson (1972) did report such a shift. To more precisely control speech rate, Robinson used timed lights as a pacing signal, and subjects counted the flashing lights aloud at a number of different rates. The instructions were to make the spoken number last as long as the light. Delays of 90, 135, 180, 270, and 360 ms were combined with masking noise. At the slower speech rates, maximal impairment (measured by errors) tended to occur at the 360 ms delay, while at the faster speech rates, maximal impairment occurred at 90 or 135 ms. Further evidence that the critical interval may shift comes from an unpublished dissertation of R.L. Frankenberger's (cited by Gates, Bradshaw, and Nettleton, 1974), which reports that short syllables are affected by short DAF values, while long syllables are not, suggesting an interaction between unit length and "critical delay".

The balance of the evidence in adults seems to be against the view that the critical interval changes with performance speed, though there is also conflicting evidence. It is worth noting that changing speech "rate" is a complex motor adjustment involving many different parameters (e.g., phonation time for vowels or nasals, length of sibilants, pauses, etc.); tapping may be a better, more controlled domain for testing this question.

The question of critical interval shift is interesting for a number of reasons. If feedback correction takes place in terms of a unit which might vary in length (for example, a long or short syllable), this would suggest that the amount of impairment might depend on the unit length, something which is contradicted by the above reports. However, a theory that proposed that feedback correction always took place within a fixed time might predict no shift. One particular theory addressed later (the Node Structure Theory of MacKay, 1987) also predicts that the critical interval shouldn't shift. Thus, this particular empirical point bears on a number of theories.

2.1.4 Extraneous Feedback: Irrelevant Speech

One important question raised in the Introduction was the extent to which DAF impairment is due to a specific relation between speaking and the delayed feedback, or is due simply to the presence of an interfering signal of some kind. Data on speech under distracting auditory stimuli (e.g., other voices) is thus relevant.

Speaking while listening to another voice seems not to be particularly impairing (at least as far as articulatory errors go; there is some impact on duration). Butler and Galloway (1957) played a subject's own voice reading numbers while they were engaged in a task reading different numbers; virtually no errors occurred when compared with synchronous feedback. Fillenbaum (1963) reported little or no effect on color name reading time when subjects heard a different voice reading color names. Yates (1965b) reports less impairment (using an omission measure) in a shadowing task when subjects hear an irrelevant voice than when they hear DAF (though there is significant impairment with the irrelevant voice compared to white noise). Perhaps of importance, Yates also found that subjects highly susceptible to DAF showed higher impairment than non-DAF-susceptible subjects when hearing an irrelevant (but not delayed) message in one channel during a shadowing task, suggesting a generalized component of auditory attention.

MacKay (1968) reported that adults' speech was slowed down about 15% when they repeated sentences while hearing another voice reading from "Alice in Wonderland", but that there were almost no articulatory errors. MacKay (1970) tested reading while subjects heard an earlier recording of themselves reading different material. Rate of speech was slightly slowed when hearing the irrelevant speech compared with a synchronous auditory feedback (SAF) condition.

2.2 Speech: Summary and Discussion

The following phenomena seem robust:

1. One major effect of DAF is articulatory errors, with the the predominant articulatory error being additions of some kind (either repetitions or insertions).

2. A second form of DAF effect seems to be overall slowdown of speech. There are two problems with this measure, however. First, there may be a speed-accuracy tradeoff, where errors occur if speed is maintained but are decreased if speed is slowed. Second, given that the predominant error is insertion, and insertions will increase elapsed time, rate and timing measures may be further confounded.

3. Such impairment in speech is highest at about 200 ms delay, decreasing with either shorter or longer delays (this figure is perhaps best assumed to apply to "young adults"). As has been frequently noted, this delay rate corresponds approximately to the average length of a syllable, and thus may be relevant to the "unit of speech control".

4. Extraneous speech is less impairing than DAF.

The results on the question of whether the delay of maximal impairment varies with actual speaking rate seem to be mixed. This question will be addressed for tapping in the current experiments.

2.3 Music

Music is another human behavior which, like speech, is directed towards auditory production, although usually with the fingers as articulators. Most studies of music and feedback have used keyboard performance, partly because data collection with electronic MIDI keyboards is fairly easy, and also because electronic keyboards provide the ability to manipulate feedback. Note that musical performances often involve complex synchronization of the two hands.

2.3.1 Feedback Absence

Feedback absence studies in music have all used keyboard. Gates and Bradshaw (1974) reported *no* impairment (with an elapsed time measure) when feedback was absent. Banton (1995) reported that sight-reading by piano players was not impaired in the absence of auditory feedback. In a study using more sophisticated measurements (Finney, 1997), I found no measurable impairment with feedback absence in measures of note errors, elapsed time, keystroke velocity, timing consistency, or inter-hand coordination. Repp (1998), however, did observe significant effects of auditory feedback deprivation on piano performance when (musically) expressive parameters were looked at; these effects are described as “very small” (except for one involving pedalling). Thus, surprisingly, auditory feedback doesn’t appear to be necessary for musical performance.

2.3.2 Delayed Feedback

Havlicek (1968) tested subjects sight-reading music with a variety of acoustic instruments (including piano); the sound was delayed by approximately 200 ms. Performance under DAF was louder than with SAF and included more errors. Gates and Bradshaw and colleagues have performed a number of studies of music and feedback manipulation, using an electronic organ. In their experiments, subjects played a practiced piece from notation “as fast as possible”, with the dependent variable being total elapsed time to perform the piece. Bradshaw, Nettleton, and Geffen (1971) reported that 200 ms delay was more disruptive than delays of 400, 550, 750, or 1100 ms in a task combining SAF with DAF. Gates and Bradshaw (1974) found that keyboard performance under 180 ms DAF took longer than with performance with synchronous feedback. Gates, Bradshaw, and Nettleton (1974), in a task combining DAF with SAF, report that of 12 delay intervals between 100 and 1000 ms, 270 ms created maximum impairment in an elapsed time measurement; this might be preliminary evidence for a critical interval in music performance. However, no controls were done on individual subject's rate of playing performance time (though a post-hoc analysis suggested that the slower and faster subjects were similarly affected).

Finney (1997) found significant impairment in musical performance with 250 ms DAF in a number of measures of keyboard performance (note errors, elapsed time, keystroke velocity, and interhand coordination): performance time was not controlled (individual subjects played at their preferred tempo). The note errors tended to be insertions (approximately 60% of the total errors in the DAF condition). Finney also reported a relative lack of impairment with pitch alterations which were synchronous with the keystrokes. An additional interesting finding was that DAF impairment was *reduced* when it was combined with pitch alteration.

2.3.3 Extraneous Feedback

Gates and Bradshaw (1974) found no significant difference in mean elapsed time between subjects playing under 180 ms DAF (which, it should be noted, is not their interval of maximal disruption) and when they heard an extraneous piece, although Gates and Bradshaw note that 10 out of 12 subjects were worse in the DAF condition. The extraneous feedback conditions was, however, more impaired than either synchronous or no feedback.

2.4 Tapping

Tapping and other rhythmic tasks were reported as impaired by DAF in the earliest DAF studies. Lee (1950b, 1951) made the observation that a Morse code operator made insertion errors under DAF, and also noted that a tympanist played arhythmically under DAF. Kalmus et al (1955) had subjects clap in groups of 6; when done under 250 ms DAF, subjects often tapped groups of 7 (that is, inserted a clap). Kalmus reports that subjects were sometimes unaware of their mistakes.

Other studies have looked in more detail at simple tapping tasks. These are the studies most directly relevant to the present research, so I will discuss them in detail. These tapping studies often had an audiometric focus (i.e., manipulating sound levels near the perceptual threshold); since my own interest is in *impairment*, I will report only the higher volume conditions (since this is where impairment occurs).

The following terminology will be relevant for the tapping discussion. "Inter-tap interval" (ITI) is the time between two taps. More precisely, I will use it here to mean the time between recorded key presses, although this is not the only use in the literature. Most DAF studies have used *grouped* patterns (e.g., tapping groups of three). In such a case, there are (at least) two classes of ITI: the ITI between taps in a group (what I will refer to as within-group interval, or WGI), and the time between the last tap of one group and the first tap of the next group (between-group interval, or BGI). Some studies have also used regular tapping, where subjects simply maintain a regular, isochronous, unstressed beat with their tapping finger.

2.4.1 Feedback Absence

Implementing auditory feedback absence is much less problematical in tapping than in speech, since there is no sound generation internal to the body. However, only two studies have addressed this question. Chase et al. (1961b) investigated the effects of masking noise on regular (isochronous) tapping; the masking served to block auditory information from physical key noise. They reported that masking noise (decreased auditory feedback) caused increased intensity and *increased* (faster) tapping rate. The statistics are a little questionable and hard to understand (for instance, although a speedup is reported, the provided median rate values are actually equal). Laszlo (1967) reported that there was no impairment with absence of auditory feedback in either fast or rhythmic tapping.

2.4.2 Delayed Feedback

Chase et al. (1959, 1961a) had subjects tap (self-paced) on a key in groups of 3 under conditions of synchronous and delayed (244 ms) auditory feedback. For tapping, DAF caused a significant increase from .37 seconds to .5 seconds in the unit-to-unit time (loosely, WGI), as well as increased loudness and some insertions (added taps), although only 2 of 12 subjects made such errors.

Chase et al. (1961b) had subjects tap under DAF in both regular and patterned (groups of 3) tasks. Subjects were trained to an ITI (WGI) of of 333 ms; the groups in the patterned task were to be separated by twice this interval, and the delayed feedback was a click presented 300 ms after the tap. Rate (what they call "unit-to-unit time", although this combined WGI and BGI in the patterned case) and intensity were used as dependent variables; errors were not counted. [3]

DAF increased keystroke force in regular (isochronous) tapping, although it did

[3]Unfortunately, timing rates are given as millimeters on a graphic pen record, and are not converted to milliseconds. Furthermore, median scores are used which, because of the method of reporting, sometimes makes it difficult to look at the magnitude or direction of the effects. In other studies reported from this period, it is often difficult to extract useful magnitudes or statistical significance.

not significantly affect tapping rate. For patterned tapping, DAF increased both intensity and WGI. Although error rates are not reported, Chase et al. (1961b, p 165) note that "there was a tendency to tap 4 times instead of 3 under all delayed sensory feedback conditions". They conclude (p 164) that "motor events which are more complex in time are more easily disturbed by DAF".

Chase et al (1961c) had subjects perform self-paced tapping of a 4 tap pattern, which consisted of a group of 3 taps, followed by a pause, followed by a single tap (what I will call a 3-1 pattern). Impairment was measured by group-to-group time, i.e., from the start of one group of 4 to the start of the next group of 4 (thus, a measure that will increase with inserted taps). DAF at very low feedback levels (10 dB above subject's threshold of hearing) caused reliable impairment relative to SAF; impaired performance occurred in 15 out of 20 subjects.

Ruhm and Cooper (1962) state that their pilot studies showed that tapping a 4-2 pattern caused more marked DAF effects than tapping such a 3-1 pattern. They had subjects tap such a 4-2 pattern on a key (blocked from view), and presented the delayed signal to only one of the two ears, while testing delays of 100, 200, 300, and 400 ms. One measure was interpattern time (the "group-to-group" time of Chase, above), which appears to be the time between the initiation of one 4-2 sequence and the next 4-2 pattern. At 5 dB above threshold, this measure was significant at all delay times, though highest at 200 ms. This is preliminary support for a critical interval in tapping, although rate is not controlled. Counting errors (here, defined as a pattern which did not contain 6 taps) at the 5 and 10 dB level were only significantly increased at the 200 ms delay. Tapping force increased for the 100 and 400 ms delay.

The fact that maximal impairment seems to occur at 200 ms (as with speech) is taken to suggest that some sort of general monitoring mechanism is at work in the two cases; one which is not dependent on syllables per se. Unfortunately, since the subjects chose their own tapping rate (and no information about subjects' individual or mean tapping rates are given), it is impossible to look at the relation between ITI and delay value. This relation will be controlled in the experiments in this dissertation.

Ruhm and Cooper (1963) presented pure tone pulses 5 dB above threshold to one ear, under SAF and 200 ms DAF conditions, again using a self-paced 4-2 pattern, with the hand blocked from view. Dependent variables were as in Ruhm and Cooper (1962), except that counting error ("number index") was changed to separately count errors in the 4 group and the 2 group. Ruhm and Cooper's data showed increased time, increased force, and increased counting errors at 5 dB DAF.

Karlovich and Graham (1966, 1967) tested the effects of DAF when patterned tapping was paced by an ongoing (visual) pacing signal. Such on-line pacing is in some ways a fundamentally different task from self-paced tapping and, perhaps not surprisingly, virtually no counting errors were found. DAF nonethess caused some temporal disruption (increase of pattern length, defined as the time between start of the first tap and end of the last tap in a pattern), and increase in tapping force.

2.4.3 Tapping: Summary

There appears to be little DAF impairment in isochronous tapping, although it is hard to apply "error" measures here (variability measures *can* be used, although such measurements have not been reported). DAF impairment seems to occur robustly in patterned tapping, with both timing and and counting errors reported as affected. In general, these studies provide little detail on error types, nor on relations between timing measures and error measures, and the timing of tapping has usually not been carefully controlled. There is limited data on the presence of a critical interval, and none on whether it changes with tapping rate.

2.5 Morse Code

Morse code is a particular form of skilled single-finger patterned tapping, in which the task involves tapping a series of temporally patterned units (letters), each made up of a sequence of 'dots' and 'dashes' (short and long key down times, respectively). This can be viewed as a very complex, temporally varied tapping task, performed under normal circumstances with synchronous auditory feedback.

2.5.1 Feedback Absence

Yates (1965a) tested 6 skilled Morse code operators in the presence of loud masking noise. Subjects made no errors, suggesting that auditory feedback is not *necessary* for the task.

2.5.2 Feedback Delay

In further experiments with professional Morse code operators under DAF, Yates (1965a) used a delay of 180 ms, and found that error rates were similar whether transmission occurred slowly, as fast as possible, or at subjects' preferred rate. Errors occurred primarily for the longer, more complex letters (requiring 3 or 4 taps); few errors occurred for shorter letters even though the length of these letters overlapped with the delay signal. Yates characterizes the errors by saying that they "almost without exception involved the transmission of an additional symbol". An experiment on critical interval tested delays between 0 and 300 ms, with subjects transmitting "as fast as possible". There was little impairment for delays from 0-100 ms (on letters of any length), but significant impairment for delays of 100-300 ms (with, however, little difference among these delays) None of the delays significantly impacted transmission rate (i.e., there was no slowdown). Subjects were described as being relatively unaware of their errors; one subject was not affected by DAF.

Howell, Powell, and Khan (1983) performed a followup experiment to Yates (1965a), with their primary concern being the absence of any obvious critical interval in the Yates study. They note that although Yates found clear error impairment with Morse code tapping, he did *not* find an effect on rate. Given the possibility of speed-accuracy tradeoffs, they chose to investigate the effects when trials were required to be error-free. Their experiment used unskilled subjects, who were trained to tap a subset of 6 Morse code letters, and then were tested on lists which were made up of sequences of two letters from this set. The dependent measure was the total elapsed time to tap 40 letters error free, and delays of 0, 100, 150, 200, 300, and 500 ms were used. Peaks of slowness occurred around 150 and 200 ms, and in 2 of the 3 letter pairs again around 500 ms; it is hard to compare the times across letters since different patterns were being tapped (though Howell et al. claim that the results suggest that amount of disruption caused by DAF is increased for more complex rhythms, ones consisting of both dots and dashes).

2.6 Discussion

2.6.1 Comparison

This section has reviewed feedback manipulations in three different domains of serial behavior: speech, tapping (including Morse code), and musical performance. In all domains, it appears as if feedback absence (or masking) is minimally impairing: it may have some effect on rate, but seems to cause only a fairly subtle effect on errors. DAF is, however, highly impairing in all these domains.[4] There is robust

[4]A particular report of shared DAF impairment between speech and tapping is Chase et al (1959, 1961a), who had subjects do two self-paced tasks, tapping on a key in groups of 3, and saying the letter ‘b’ in groups of 3, under conditions of synchronous and delayed (244 ms) auditory feedback. Similar impairment (e.g.,

evidence for a critical interval (a delay value causing a peak of impairment) in speech, and tentative evidence for one in music, although the actual value may be different, and rate has not been carefully controlled. Insertions seem to be the most prominent error in all domains, though detailed information has been provided only for speech. Based on the speech and music data, extraneous distracting stimuli seem less impairing than DAF.

The similarities in impairment across these domains provide provisional motivation to think the same mechanism may be active in DAF impairment in all 3 domains. Thus, tapping can be chosen as an area of study, and the results might tentatively be expected to apply more generally.

2.6.2 Empirical Questions

This dissertation will address three particular empirical areas suggested by the above review. One task here is to look at areas which have given mixed results in speech and see what the results are for tapping (e.g., critical interval shift). A second task is to look with more precision at the impairment in tapping, with one goal being to see whether additional similarities (or differences) to speech will be revealed.

1. Determine in more detail just what the effects of feedback *absence* are (if any) for tapping.

2. Investigate in detail the critical interval in tapping. This involves determining more precisely whether there *is* a critical interval, and if so, whether it shifts

---

increased loudness, increased time between units, and some insertions) occurred in both cases. Chase et al. take this similarity in results as indicating that the same process operates in vocal and non-vocal tasks. It is important to note that the vocal task was *not* continuous speech, but rather a discrete counting task, so it is unclear if these results should be taken as generalizing to normal speech. Also, Chase et al. (1961a) found little correlation *within a subject* on these two measures, which seems problematic.

with tapping rate.

3. Look in more detail at the impairment pattern in tapping. One particular issue is to look at insertion rates, and see whether the report that the relative proportion of insertions is highest at the critical interval (reported for speech by Fairbanks and Guttman, 1958) will also hold for tapping.

Chapter 3

Theoretical Review

In this section, I will discuss the proposals that have been put forth to explain the impairment found under DAF. Most of these proposals derive from the research in the speech domain (since that is where most of the DAF work has taken place), but many can be extended in obvious ways to non-speech domains, and some are explicitly intended to have wider application. One point here is to establish which existing theories are worth looking at further (with respect to both predictive power and empirical adequacy), and to establish interesting predictions to be tested in the tapping domain.

Existing proposals addressing DAF can be loosely classified into two categories: those that describe certain factors of the experimental situation which are linked to impairment (i.e., effectively a correlational approach), and those that try to explain some of the particular symptoms of DAF impairment. I will discuss these in separate sections, but I first provide an overview of the sources of information which may be relevant in the DAF situation.

3.1 Information Sources

In evaluating these theories in any detail, it is worth keeping in mind the number of sources of information which are involved in any motor task. Feedback is not a monolithic "sensory event": many different, perhaps independent channels may be involved in a given action. Some hypotheses later will use the presence of some discrepancy among different feedback (information) sources as part of an impairment explanation. The information sources are the following:

1. There is presumably some sort of cognitive "motor plan", which issues commands to lower-level systems; a copy of the "motor plan", or an efference copy of the motor impulses, may be kept in memory.

2. There is *feedback* from the action: perceived sensory consequences of the action. Within this broad area, there are a number of somewhat independent channels, depending on the action: e.g., auditory, visual, kinesthetic, tactile, and proprioceptive. The information from the different channels may be of different forms (e.g., sound and touch), although they will presumably be correlated for a given action. In the speech case, there is the interesting factor of separate bone and air conducted feedback (the latter being the one affected by DAF), where there are basically two different channels carrying the "same" auditory information, albeit (in the DAF case) with a phase shift. [1]

3. There may be an image of the desired result of the action; such an image could conceivably exist in any of the modalities listed in (2). Note, interestingly, that

---

[1]Note that there are likely to be differences in frequency between bone and air-conducted feedback (Howell and Powell, 1984) as well as small time differences based on the relative speed of bone and air conduction. Thus, there is already at least a *small* phase shift.

in the ideo-motor theory (see, e.g, James, 1890) such an image might actually be the stimulus for action.

How might different channels be compared to ascertain that there is a discrepancy? In the case of bone and air-conducted auditory feedback, the two channels at least carry (approximately) the same sort of information. However, comparison of other, less compatible channels (e.g., auditory and tactile) may be a much more complicated problem, since the types of information are quite different. Note also that any form of error correction or feedback guidance requires having some sort of standard against which to compare the percept; the modality and character of this standard also needs to be specified. Looking at information sources in detail thus requires being aware of the "units" that each is processed in, and the content and comparability of those units (loosely, the "vocabulary" of that channel).

Some of the feedback channels listed above can be used unconsciously and automatically to correct action (e.g., certain types of muscle control may rely on proprioceptive feedback, e.g., adjusting arm muscles for varying weight loads).[2] There are also clear examples of feedback use for error correction, for example, at least some cases of error correction in typing, music and speech.

3.2 Theories: Descriptive

The theories in this section aim at describing certain characteristics of the DAF situation which may be associated with impairment, but do not attempt to actually explain the process of impairment. They may, however, suggest areas that should

[2]Note that feedback correction is not the only approach to this; the dynamical systems approach (see, e.g., Kelso, 1995) would postulate that such a balance/correction is an emergent property of coupled or competing systems. Although such explanations are plausible for simple behaviors, they are hard to extend to complex serial behaviors such as speech.

be looked at more closely; i.e., if a correlation truly holds, it may at least suggest where to look for the sources of impairment.

3.2.1 Feedback conflict

Smith (1962, p 55), Ruhm and Cooper (1962), Yates (1963a, p 228), Havlicek (1968) and Borden (1979) all suggest that impairment in DAF is due to a conflict or mismatch between the feedback coming in from different channels. The channels involved are not always precisely specified (perhaps delayed auditory and synchronous (bone-conducted) auditory, or perhaps delayed auditory and undelayed proprioceptive). The main problem with such proposals is that they are seriously underspecified. First, it is basically a description of circumstances causing impairment, without any explanation of the process by which feedback conflict causes the behavioral consequences. Second, it avoids the very interesting "binding" problem of how the information from different feedback channels is viewed as associated or disassociated (normally, temporal synchrony probably *is* the means by which this occurs). Similar problems also exist in feedback control theories, discussed later.

3.2.2 Overlap

Related to the idea of feedback "conflict" is the extent to which impairment occurs when feedback from a prior event "overlaps" the next event. One specific mention of this is Chase and Guilfoyle (1962), who note that when impairment occurs it often involves overlap between the synchronous auditory feedback for a "vocalization" (a very unspecific unit choice!) and the delayed portion, and that this is one source of DAF impairment. Chase et al. (1961b) suggest that the unusual overlap of the delayed signal with the existing feedback might cause distortion of the correct feedback; should there be no overlap, there should be no impairment (this solution is consistent with lack of impairment with feedback absence). Note, however, that this proposal implicitly contains reference to some (unspecified) unit in which the feedback overlap occurs. Chase et al. suggest that tapping studies (where overlap can be controlled) would be a good area for investigation of this question.

Another form of (speech-specific) overlap proposal is provided by Zimmerman et al (1988), who provide details of the relative timing of jaw movements in different conditions of auditory feedback (synchronous, and DAF of 100 and 200 ms). They analyzed the articulatory errors in their data set, and report that repetitions occurred when the DAF (which they view as coming in discrete units of voiced vowels, separated by pauses caused by consonantal closures) overlaps the following CV gesture. Such overlaps (due to the timing of speech, and relating to syllable length) are more likely to occur at 200 ms delay than 100 ms, and thus the critical interval. However, they only had 7 instances of such errors to analyze, and many such overlaps did *not* cause repetitions.

These two proposals involve different forms of overlap. For Chase, it is overlap of delayed feedback with undelayed feedback from the same "unit" (why, then, the critical interval, since maximal overlap will occur with shorter delays?), while for Zimmerman, the relevant overlap is feedback from a previous unit with a subsequent "gesture".

One problem with either overlap approach is that they don't readily account for the (relative) lack of impairment effect when there is irrelevant speech (which would certainly cause *some* form of overlap). Thus, such theories may require awareness of internal generation, or of particular unit identification. But the main problem with such proposals is that, again, they are simply descriptive, and they need elaboration to be a genuine theory of DAF impairment.

3.2.3 Displaced Rhythm Hypothesis (DRH)

A somewhat more precisely specified descriptive theory is the "Displaced Rhythm Hypothesis" (DRH) of Howell, Powell, and Khan (1983) (see also Howell, 1984; Howell and Archer, 1984; Howell and Powell, 1984, 1987). This proposal has two components: one is that only the timing but not the identity of the feedback units is important, and the other is a (partial) specification of the particular timing relationship which causes impairment. Howell et al. thus argue that DAF impairment has nothing to do with use of auditory feedback for control, but is a form of interference.

The particular rhythmic signal which causes impairment is argued to be one in which one signal goes on just as another signal is going off, e.g., Howell et al. (1983, p 774) describe the situation as follows: "...syllables are about 200 ms long and when speech is delayed by 200 ms, the amplitude of the spoken syllable would be falling to its minimum when the amplitude of the displaced signal begins to rise to its maximum...at delays less than or greater than 200 ms, the intruding rhythm would overlap with a syllable that is being issued so that the rise in amplitude of the displaced signal would be less intrusive...thus, DAF at a delay of 200 ms would create a particularly intrusive distracting rhythm rather than distort feedback, as proposed by earlier theories." Thus, for speech, the discrete signals are defined by the amplitude contours of syllables.

The experimental evidence they provide relies on an elapsed time measure of error-free trials for read vowel lists, and the emphasis is the critical interval near 200 ms. Their support for the DRH primarily comes from studies in which the signal contents is altered (so unit identity information is lost), but DAF impairment still occurs. This both argues against unit identity as important, and suggests that it is only the rhythm which matters. Howell, Powell, and Khan (1983) delay a significantly distorted version of the speech signal, and find impairment similar to DAF.

Howell and Archer (1984) gate a square wave (clearly containing no useful phonetic information) with a delayed amplitude envelope derived from the produced speech; impairment at 100, 200, 300, and 400 ms is described as being indistinguishable from that of DAF. Howell and Powell (1984) used the DAF amplitude to gate a sawtooth wave; a peak in impairment was found at 200 ms delay, similar to speech DAF. Howell and Powell (1987) use the delayed speech envelope to gate a sawtooth wave, and find a similar pattern of impairment as when speech is delayed (using delays of 0, 100, 150, 200, 250, and 300 ms, peak time impairment occurred at 150-200 ms in both cases, and to a similar extent in the speech and non-speech cases). Such results argue against the idea that feedback *units* are relevant.

Timing Relationship

To evaluate this theory for tapping, it is necessary to look in more detail about what the exact timing relationships are. Howell et al. (1983) are in fact somewhat ambiguous about this, but the quote above (as well as their diagrams) suggest that the offset/onset relationship must be quite close. It is worth looking at existing speech data to determine what timing relations must occur if the DRH is to hold. Although their data deals with isolated vowels, maximal impairment occurs (as it does with connected speech) at about 200 ms delay. The range of 150-250 ms is the typical estimate for syllable length (although whether this applies to isolated vowels is perhaps open to question). Combining these two numbers suggests that the onset of the delay signal must occur within about 50 ms of the offset of the spoken syllable for maximal impairment to occur. A-priori, one would assume the same timing relationship should apply to tapping.

Critique

There are a number of issues with the Delayed Rhythm Hypothesis. The first is that, once again, that even if the hypothesis is true, it is a description, not an explanation. No description of the impairment *process* is provided.

A second concern involves vagueness about just what the two signals entering into the impairing relationship are. Howell (1983) seems to explicitly view the impairment in terms of two conflicting auditory signals (for speech, the bone-conducted and delayed air-conducted auditory feedback), though Howell et al. (1983) note that the perceptual events need not be in the same modality, and further refer to the general case of "two rhythmic events". Thus, for tapping (which only has one auditory channel) the alternate channel providing conflict would need to be another modality (e.g., proprioceptive). An alternative is that the signal conflicts with the motor act itself; e.g., Howell and Archer (1984) state that disruption occurs "when a rhythmic event goes on concurrent with but out of synchrony with the *activity*" (emphasis added). Howell (1983, p 394) explicitly notes that DAF disruption in non-speech activities would be explained by the DRH if "a similar relationship between the *activity's* offset and the onset of feedback can be identified" (emphasis added). For tapping, the most obvious interpretation (and the one that will be followed here) is one in which key *release* coincides with the *onset* of a feedback signal.

Howell et al. also provide little detail about the precise constraints on the timing relation between signal off and (disrupting) signal on, a relationship which also must be made specific for any predictive power. (Detailed timing information on the actual articulations and placement of delay signals in the data is never provided.) Also, the use of a simple task (vowel lists) with a crude measure (elapsed time) raises concerns about the application to more complex stimuli (continuous speech) and error measurments.

Thus, the DRH makes the prediction for tapping that impairment will occur when feedback tones start very close to the end of a tap (that is, key release).

3.3 Theories: Explanatory

Some other theories more specifically attempt to explain the process behind some symptom of DAF impairment in the content of more general principles. Here, I will focus on theories which explicitly address insertions,

3.3.1 Serial Chaining

Serial chaining theories of sequencing assume that one unit in a sequence is the stimulus for execution of the next unit. In the feedback-based form of serial chaining, the feedback from one executed event is the actual stimulus-response trigger for the start of the next event in the sequence. Lashley (1951) claimed that rapid performance (e.g., piano playing with 15-20 keystrokes a second) was too rapid to be implemented by such a feedback-based chaining hypothesis. There are two problems with accepting this particular dismissal of chaining theories. First, even if true, it leaves the possibility that (more normal) slower acts might be implemented in this way. Second, the time scales actually involved in feedback response are somewhat controversial. Although studies such as Keele and Posner (1968) have provided data suggesting that visual information is not used for correcting tracking unless it occurs at least 190 ms before the act, other studies have suggested that the feedback reaction time may be much less than that (and, of course, may vary across modality and task). Carlton (1981) suggests the time for visual correction of action is closer to 135 ms. MacKenzie and Marteniuk (1985) review studies suggesting that feedback correction may be active in some circumstances at latencies of as little as 100 or 150 ms. Bunker (1912) does not address this issue. Bootsma (1990) provides a figure for visuo-motor delay of between 100 and 120 ms, and cites other studies that claim it might be as low as 50 ms. A figure of 100 ms would still limit the number of actions to about 10 per second (in a single "chain", at least), but this covers much of skilled human behavior.

One particular serial chaining theory for insertions in speech DAF is due to Fairbanks and Guttman (1958). In their theory, feedback from one unit is the cue for execution of the next response, and DAF interferes with the normal processing of such sequences. Some details of their proposal are unclear, so I will provide two different possible interpretations. In the first interpretation, execution of a second unit cannot proceed until feedback from the first is received. In this case, execution of the first might perseverate until the second is executed, and this perseveration might manifest itself as a repetition. This would seem to predict that every single unit would be repeated. A second interpretation (and the one perhaps intended by Fairbanks and Guttman) is that the second unit executes without auditory feedback (perhaps responding to feedback from a different channel), but that delayed appearance of the auditory feedback from the first unit may cause a *second* triggering of that second unit if the auditory feedback "dominates the feedback complex during the second response". The explanation for exactly when repetition or proper sequential cueing does or does not occur then apparently depends on competition among feedback channels. They describe this as an automatic response to misinformation in the feedback complex, rather than error correction.

There are two major issues with such serial chaining proposals. The first is that the presence of feedback is central to sequence execution, but, as discussed in Chapter 2, auditory feedback absence does not cause major impairment in a number of serial tasks. Although other feedback channels may then be used (as seems implicit in the proposal of Fairbanks and Guttman), the question then arises of why DAF is impairing. A second issue was noted by Lashley (1951): such an approach has difficulties accounting for the flexibility of sequencing in motor serial chaining theories, where typically any given unit can be followed or preceded by many other different units. Thus, it is necessary to explain how the associations are created, since they are not permanently fixed.

3.3.2 Feedback Guidance

"Feedback guidance" or "error correction" covers a wide range of proposals. This section will describe proposals relevant to DAF, again with a focus on insertions. These theories typically *assume* that feedback guidance is part of normal performance, and then discuss precisely how it is implemented.

The basic idea behind feedback guidance theories is that of feedback control theory (see, e.g., Toates, 1975; Wiener, 1948): a sensory consequence of an action serves as the source for generation of an error signal. That is, the difference between the actual output and the *desired* output (which must be stored somewhere in the system) is compared, and the difference is used to take corrective action to bring the system closer to the desired goal. Such a theory is most typically applied to continuous, quantitatively measurable tasks, such as the homeostatic maintenance of a home thermostat, or tracking tasks (such as the gunnery tasks which originally motivated control theory).

Feedback control in language (and serial behavior in general), however, is not dedicated to homeostasis or following an external stimulus but involves controlling sequential, pre-planned, self-generated output over time. Attempting to apply feedback theories to this domain raises a number of issues:

1. If feedback is used to correct performance, a crucial question is what units this is done in. For instance, speech can be viewed as sequences of phonemes, syllables, or words (among other possibles); which of these are subject to

correction?

2. For tapping, feedback control must be applied to discrete units (taps, or perhaps groups of taps). That is, at a certain level a tap is either executed, or it is not executed; it cannot be "a little bit off". Nor can a tap, once it has been excuted, be corrected (unlike, say, a tracking task).

3. An interesting and foundational issue is that of how is an association made between an (error) percept and the act which produced it (or the template which is the standard for "correct"). This is a non-trivial binding problem (at least if there is any temporal discrepancy), and one which is not often addressed There may be a question of the comparability and the vocabulary of the two signals: comparing sound with sound is presumably (relatively) straightforward, but how do can sound be compared with a motor program? What channels are chosen for correction, and how are they combined?

The following sections will describe some particular feedback proposals addressing DAF.

Lee

Lee (1950a) views the stuttering effect of DAF as a reflex act at the syllable level, where the lack of successful "monitoring" of a syllable causes it to be repeated. Such a theory is consistent with the semi-match between delay of maximal impairment and length of a syllable (if one assumes that failure will be recognized at the *end* of a unit).

Such a simple proposal can also be applied to tapping, where a tap might be repeated if feedback for it is not received in a timely manner.

Fairbanks

Fairbanks (1954), along with Fairbanks (1955) and Fairbanks and Guttman (1958), outline an explicitly "cybernetic" proposal for a feedback-based system of speech control. However, as described in Section 3.3.1, the approach they take to accounting for repetitions under DAF is actually much closer to a serial-chaining proposal than a feedback correction proposal.

Fairbanks notes that when speakers hear their own output, this can be be used to monitor the success of speaking. In order to associate the sensory information with a stored desired or correct version, one possibility is that there is only one unit active at a time. The comparison of feedback with this stored reference signal is used to adjust the motor system to make the two match. The next unit is triggered only when the error signal has reached an acceptably low level. Such matching may occur in multiple sensory dimensions (auditory, proprioceptive).

However, the feedback system does not merely try to get the error signal to zero, but is actually designed to control future output. To help account for continuous performance, rather than waiting for the "error signal" to reach zero, continuous predictions are made as to when (in the future) it *will* reach zero, and output of the next unit commences when this occurs. This seems to implicitly assume a continuous control signal (i.e, a comparison to zero, and presumably the magnitude of change is continually dependent on the magnitude of the error signal).

Fairbanks purposefully avoids making any strong claims about the identity of the unit of control (although suggesting that it is *not* a standard linguistic unit such as the word or syllable, but a " semi-periodic, relatively long articulatory cycle". He explicitly states that DAF research may contribute to understanding the unit of speech control. Fairbanks (1954) states that this model can account for DAF articulatory effects (including repetitions) but, as noted, the actual explanation in

Fairbanks and Guttman (1958) doesn't really seem to be an error-correction theory.

Chase et al.

Chase's work (1959, 1961a, 1961b) makes a number of different, somewhat underdeveloped observations about the DAF situation. Two suggestions in particular, however, might be part of a feedback theory for tapping.

Chase et al. (1961b) suggest that DAF impairment occurs because the delayed feedback is taken as accurate information about motor performance, and corrective action is taken. Chase et al. (1959, 1961a) argue that the critical delay of 200 or so ms suggests that feedback must be received within this period for optimal motor control.

For tapping, these could be made part of a simple feedback control theory as follows: when auditory feedback is present, it is used as confirmation of tap success on a per-tap level. There is a fixed period of time (say, 200 ms) within which such feedback must arrive; if feedback doesn't arrive, the tap is repeated.

Huggins (1968)

Huggins (1968) assumes a sequential control program with a marker for the current place in the program. When place in the program is lost for some reason, the current articulation continues until the place is refound. If the place can be rapidly found by bone feedback or proprioceptive feedback, the only effect is a slight lengthening; if the place is found by use of loud, salient delayed feedback, then that segment may be repeated (since the program is restarted from an earlier location). Any use of the auditory information to find the correct place in the program is dependent on the program and the auditory information being "similar". Such a theory is harder to apply to tapping, which has discrete and non-identifiable units.

Feedback Theories: Discussion

In general, these theories are somewhat underspecified. Also, most of these theories predict increasing impairment at higher delays, which does not match the report of an peak of impairment.

An additional major issue with any theory positing that (auditory) feedback guidance is a normal part of serial motor performance (speech, music, or tapping) is the fact that the evidence reviewed in Chapter 2 suggests that in all of these domains auditory feedback absence is impairing at a very subtle level, if at all. It certainly doesn't cause the gross impairment of DAF. However, *mismatched* feedback could still have an effect. For instance, auditory feedback might not be necessary, but may be obligatorily used if it is present. Alternatively, it could be that the system only activates when there's a signal that an error has occurred (and absence of feedback might not classify as an error signal), although this approach cannot apply if the feedback is explicitly used to confirm that the act occurred, rather than just whether some aspect of the act needs correction. Note that feedback (perhaps even separate systems) could be used for (1) fine continuous control and (2) "gross" error correction, without being necessary for basic order, timing, and sequencing.

The idea that the DAF effect is indicative of malfunctioning of a normal feedback system is far from universal. Borden (1979) suggests that the DAF impairment is not due to any normal role of feedback guidance, but to the fact that the abnormal DAF attracts "attention", and invokes an unusual corrective process Lane and Tranel (1971) also argue that the DAF effect is an abnormality, and not indicative of a significant role for auditory feedback in normal speech (rather, it is an attempt to maintain communicative function).

3.3.3 Node Structure Theory (NST)

MacKay (1987) proposes a general purpose, wide-ranging theory of perception and action, with one explicit goal being an explanation of the DAF effect. I will attempt to provide an explanation sufficient to show how his explanation of DAF works, but will not discuss the entire theory.

MacKay's "Node Structure Theory" (NST) is intended to account for both the similarities and the differences between perception and action. A crucial concept for this is the "content node", where, crucially, content nodes are *shared* by both action and perception. Activation of a given content node can (depending on other aspects of the system) either cause an action, or give rise to a percept. As an example, the node for the word "theory" will be activated relatively early in the process of speaking the word "theory" (which will, in turn, cause activation of the lower level motor system nodes which actually cause the utterance); similarly, during perception the same node will be activated as a consequence of priming from (the sequence of) lower-level perceptual nodes corresponding to the sounds for 'th', 'e', etc.

Priming (a necessary precursor to activation) spreads automatically between linked nodes, and, in order for nodes to be shared between perception and action, it must in the general case be able to spread both top-down (e.g., from the word node to lower level action nodes) and bottom up (e.g., from lower level acoustic nodes to the word node). This bi-directional activation is very general: if a node primes a lower level node (which then becomes activated), that lower level node will in turn send priming back to the original, higher level node. Such a process might cause repeated activation of the original node (since nodes maintain activation for some period of time), and repeated action when only one action was intended. To prevent this, nodes have a refractory (self-inhibition) period following activation, that is, a period during which the node cannot become reactivated. The refractory period also solves another problem which might cause repeated activation of the content node: feedback returning from the action (e.g., the sounds produced in speech) will also provide additional priming (after a short period) to the same node which produced them. Thus, the refractory period is also a means of preventing repeated activation caused by feedback.

The final piece of MacKay's account of DAF is that this refractory period is followed by a period of hyper-excitability before the node returns to its resting state. This entire sequence of activation, refractory period, and hyperexciteabity is termed the "recovery cycle". Although different nodes (e.g., at different levels of a content node hierarchy) may have different recovery cycles, for phonological nodes hyperexciteability is described as peaking at 200 ms, and being back to normal at 300 ms (though the cycle may also be affected by the node's history of activation).

The stage is now set for MacKay's explanation of DAF impairment (and insertions in particular). Recall that the hyperexciteability period for phonological nodes starts around 200 ms. Recall also that the delay of maximal impairment for speech is also near 200 ms. What happens is that the delayed feedback circumvents the safety mechanism of the refractory period. DAF (being delayed) shows up after the refractory period, but *during* the period of hyper-exciteability. Such reinforcement combined with the existing activation makes it likely that the node will be reactivated, causing the demonstrated repetitions under DAF.

In this theory, then, DAF impairment is *not* a consequence of feedback being used for guidance. Rather, the sound is viewed as a direct stimulus to the repetition.

Such a theory explains the critical interval, since the delay causing (maximal) impairment must be closely tied to the (relatively fixed) refractory period of the phonological node (it is also consistent with the report of MacKay, 1968, that the maximal impairment delay is not affected by speech rate). The particular common impairment syndrome of repetition found in DAF speech (as well as tapping) is also explained. Age differences in the delay of maximal impairment are explained because the timing of the self-inhibition cycle can vary across subjects (particularly with respect to age); this might also (to some extent) explain individual differences (although it would suggest that different subjects would have different peak impairments). Thus, the theory has an impressive domain of successes.

Time Course

Later experiments will look at how well the Node Structure Theory applies to tapping, by looking at the actual timing relations between inserted taps and the feedback which precedes them. However, to make firm timing predictions it is necessary to be precise about what the elapsed time is between arrival of the feedback and reactivation of the node, something left unspecified by MacKay.

Although MacKay does not specify this time interval, it is possible to look at the speech data and deduce what such a time must be. To evaluate this theory it is necessary to know what the time is between an initial utterance and its repetition. For MacKay's theory to work, this will need to be greater than the delay (thus, greater than 200 ms or so in the most impairing case).

A search of the speech DAF literature reveals 3 sources of data for this. Figure 4.12b of Kozhenikov and Chistovich (1965) provides a detailed diagram of a number of repetitions of one utterance of a Russian speaker under 210 ms delay; measurement of the diagram shows that the repetitions occur in the range of 130-185 ms after the initial utterance. Figure 4 of Zimmerman et al. (1988) provides detailed timing analyses of 3 repetitions in 2 English utterances; measurement seem to show 3 repetitions occurring at 132, 200, and 800 ms. Finally, Huggins (1968), in a detailed articulatory analysis of a single subject, reports that the time between the initial utterance and its repetition is approximately equal to the delay itself. Thus, in the high impairment condition with 200 ms delay, the time until repetition would also be 200 ms. Fairbanks and Guttman (1958) also report that repeated portions are longer with longer delays.

Thus, for the returning sound to be the stimulus for the repetition (as postulated by MacKay's theory), and taking the repetition time under 200 ms delay to be approximately 200 ms (as reported by Huggins), the effect of feedback needs to be virtually instantaneous for the theory to apply. Although such instantaneous response is not intuitively plausible, there are no studies in speech which explicitly look at data involving the feedback signal and the speech signal, so there is not direct conflicting evidence. Also, the discussion in 3.3.1 showed that reaction times to feedback may in fact be quite rapid (under 100 ms), and in MacKay's theory the feedback is input to a highly activated system, Thus the Node Structure Theory explanation of DAF may still be a possibility if one allows for a certain amount of variability in impairing delays and syllable length.

The crucial fact for the later tapping investigation is this: although MacKay (1987) does not describe a particular duration for the time from feedback to repetition, the speech data shows that this time must be close to zero for the theory to work. Since his theory is intended to be quite general, the same figure should apply to tapping.

Critique and Predictions

In this theory, the precise timing relation between length of delay and refractory-period characteristics of a node is crucial. Also, although not specifically a problem for the DAF explanation, a non-trivial question is exactly what perceptual feedback is taken to activate the "same" node as a motor impulse (this is related to the "association" question raised above under feedback).

The NST makes the following strong predictions:

1. The maximal impairment delay should not change with actual performance speed. As noted in Section 2.1.3, there are conflicting results for speech; the current experiments will address this question for tapping.

2. Insertions should be preceded by stimulus events, and, based on the speech data, the response time between feedback and insertion should be almost instantaneous.

3.4 Summary and Discussion

The following theories from the preceding discussion will be tested in the tapping domain.

1. Feedback guidance. There are many possibilities for feedback guidance theories. For present purposes, I will view the tapping task as discrete (i.e., a tap either occurs or it does not), although rhythmic tapping also has an oscillatory component (Vaughn et al., 1998). The simple theory to be tested here is the following: when a tap command is issued, the system waits a certain fixed time for feedback from the tap; if feedback is not received, the system reexecutes the tap.

2. MacKay's Node Structure Theory (NST), described above.

3. Howell and Powell's Displaced Rhythm Hypothesis (DRH), described above.

Some particular predictions made by these theories are the following:

1. The NST explicitly predicts that an insertion will occur *after* relevant feedback is processed. If the theory is correct, it should be possible to look at the extra taps which occur under DAF, and find the feedback event which stimulated

it. As noted above, for speech the relationship between feedback onset and stimulated tap must be very rapid, almost immediate.

2. The DRH also makes a specific timing prediction for the relation between taps and feedback. For speech, the delayed signal (syllable onset) must go on almost exactly as the undelayed signal is going off. Again, this precise timing relation can be looked at in tapping. If the relation holds, this would be support for the DRH. If the relation *doesn't* hold, this part of the DRH cannot be correct.

3. Both MacKay's NST and the feedback guidance theory outlined above suggest that the critical interval is fixed, and will not vary with rate. Is this true? Note that with tapping, rate can can be more carefully and precisely controlled than with speech.

These will be addressed with simple tapping tasks, a task which (at a first approximation) causes the same impairment with DAF as more complex tasks, but which is more amenable to precise controls (e.g., rate changes).

However, addressing the detailed timing questions (particularly the interaction between rate and critical interval) requires a method change from the self-paced tapping done in studies such as Chase et al. (1959, 1961a). The method used will be a continuation paradigm, in which all trials start out with subjects tapping to a precisely-timed pacing tone, and then continue on their own. Experiment 1 will therefore be a preliminary experiment which investigates the extent to which the use of this method does or does not replicate previous tapping results with DAF.

Chapter 4

Experiment 1: Preliminaries

Previous experiments in tapping and DAF have not precisely controlled tapping rate, something which is a prerequisite to a study of the critical interval. A continuation paradigm is one means for controlling rate. In a continuation tapping task (see, e.g, Fraisse and Voillaume, 1971; Wing and Kristofferson, 1973), subjects start by tapping along with pacing tones presented over headphones. After a period of this "synchronization phase", the pacing tones stop, and subjects continue tapping in the same manner (the "continuation phase") until the trial ends. Such a procedure will at least initially control tapping rate across trials and subjects. Experiment 1 differs in two fundamental ways from previous continuation experiments. First, the sequences here usually involve a gapped pattern (e.g., specifically a 4-2 pattern) rather than the isochronous sequences used in studies such as Wing and Kristofferson (1973). Second, rather than having subjects either get no auditory feedback from their taps, or synchronous feedback, subjects received no feedback during the synchronization phase, and usually received altered (delayed) feedback during the continuation phase.

Experiment 1 was designed to verify the extent to which such a continuation task would replicate existing tapping DAF results in the literature. Experiment 1 also provided additional detail on tapping in the absence of useful auditory feedback ("feedback absence), something which is sketchy for tapping in the literature, and also provided an opportunity to do more fine-grained temporal and error analyses than in previous tapping studies

Experiment 1 involved two different types of tapping (regular (isochronous) tapping, and patterned tapping), each performed at two different tapping rates. This experiment (and all succeeding experiments) was run as a single-session within-subjects design, with all subjects performing in all delay conditions. It was felt that an hour was the maximal feasible length for this type of experiment; this constrained the number of trials and conditions in an experiment.

4.1 Method

4.1.1 Equipment

The tapping device used was a Fatar 49 MIDI keyboard controller (i.e., an electronic musical keyboard); this keyboard is velocity sensitive and provides a value for the velocity of each key press. The keys are not weighted, and do not provide much physical resistance. Key travel was about 1/2", and a key press caused a signal to be sent at about the halfway point on the downstroke; key release was signalled when the key almost back to the top position. Subjects tapped on a single key; the other keys were blocked by a thin piece of cardboard. View of the tapping hand was blocked by a large horizontal piece of cardboard about 5 inches above the keyboard, and the subject's forearm was supported by a padded, 1 1/2" high wrist rest, though subjects were free to adjust this to their comfort.

The MIDI output of the keyboard was connected to an SGI Indigo computer via an Opcode MIDI Translator-II interface. A custom program using the SGI

MIDI interface recorded all incoming data (key down time, key up time, note, and key velocity), and also sent MIDI output data to a Yamaha TX-81Z digital tone generator (which was programmed to generated pure tones). The output could be modified in pitch or time from the original keystrokes. The program also generated timed MIDI signals for the pacing tones. Subjects listened to all sounds through Koss PRO/4XTC headphones.

Two different tone pulses were used for the pacing and feedback tones; they were of different pitches and different lengths. Pacing (metronome) tones were the note D7 (about 2350 Hz), and were 30 ms long. Feedback tones were the note E6 (about 1320 Hz), and were a fixed 100 ms long; the pitch, volume, and note length were not dependent on the subject's keystrokes. Volume levels were set to levels that allowed for feedback and pacing tones to be heard simultaneously; the resulting pacing tones were 79 dB, and feedback tones (lower pitched, but longer) were 78 dB. Sound levels were calibrated with a Radio Shack sound-level meter held 1/4" away from the headphone speaker.

To block physical key noise (which would provide synchronous auditory feedback information), white noise (generated by a minor modification for the TX81Z of the white noise voice in Massey, 1987) was presented over the headphones throughout each trial. In fact, the relatively high white noise level (79 dB) was not enough to stop a determined subject from tapping very hard and hearing the physical tap through the headphones. Occasionally when a subject was tapping exceedingly hard (e.g., when they admitted being able to hear themselves when questioned) they were asked to tap more softly; this occurred only rarely.[1]

---

[1]This instruction, and the resulting variable treatment of different subjects, seemed preferable to having the confound of obvious SAF along with the DAF. It compromises the use of the key velocity data, and this might be a (small) part of the reason why the velocity data were not significant in this experiment (see later

4.1.2 Procedure

Subjects were seated in front of the keyboard, and given written instructions (provided in Appendix A) that told them to tap with the index finger of their (self-reported) dominant hand on a particular white key of the keyboard, to tap from the wrist, and not to lift their finger from the key. Nothing about the pattern type (the numbers 4 or 2) or counting was mentioned in the instructions, to try to avoid biasing subjects towards a particular method of generating the pattern. The written instructions also explicitly suggested that subjects should ignore the sounds if they found them disruptive.

An experimental trial consisted of a synchronization phase and a continuation phase. During the synchronization phase, the subject heard (and tapped along with) the patterned pacing tones, and there was no auditory feedback to the user's keystrokes.[2] After the designated number of synchronization patterns (see below), the pacing stopped and auditory feedback (synchronous or delayed) to the user's keystrokes started (except in the no feedback condition). Masking white noise was on from the very start of the trial, and automatically terminated after the trial durations listed below; this was the subject's cue that the trial had ended.

Each block (tapping condition) of trials was preceded by three practice sequences. First, subjects tapped along with the pacing tone, hearing *synchronous* feedback,

---

discussion).

[2]The choice of no feedback during synchronization is somewhat arbitrary. This choice was made to avoid potential problems with "auditory imagery", e.g. the user being disrupted by the change of pattern in the sound of their taps in the DAF conditions. However, it also turns out out to be involved in a possible timing confound which compromises the timing data in the SAF condition; see later discussion and Appendix B. To make a virtue out of it, there is certainly disruption at the transition point from synchronization to continuation in the DAF trials, so it may be good to have a small corresponding disruption in the synchronous trials as well.

until the experimenter felt they were comfortable with the task (typically, 25 to 50 seconds, depending on condition difficulty). Next, subjects tapped along with the pacing tone without auditory feedback. Finally, they ran one practice trial, which was identical to a test trial in the synchronous feedback condition.

The experimenter (the author) was present throughout the experiment, and initiated each succeeding trial when the subject indicated readiness; subjects could take breaks at any point simply by not cueing the experimenter for the next trial. Short breaks were also enforced between blocks, and sometimes in the middle of long tapping blocks.

4.1.3 Subjects

16 subjects were run in Experiment 1; they received either payment or course credit for their participation. One subject reported himself as falling asleep, and there were observable gaps in the resulting data; he was removed and replaced.

4.1.4 Conditions/Stimuli

Regular (isochronous) tapping was performed at 2 different rates (250 and 400 ms); these conditions will be referred to as the "250r" and "400r" conditions (where the number refers to the rate in milliseconds, and the "r" refers to r(egular) tapping. For these conditions, there was 15 beats of synchronization, and 40 beats of continuation, giving a total time for a 250r trial of about 15 seconds, and for a 400r trial of 23 seconds.

The patterned tapping consisted of 4 taps followed by 2 taps, which was repeated a number of times; this will be referred to as a 4-2 pattern. This pattern has been used in prior studies (e.g., Ruhm and Cooper, 1962; Karlovich and Graham, 1967, 1968), which have reported that it showed more marked effects than simpler patterns such as groups of 3, or the pattern 3-1.

The same two rates (250 and 400 ms) were used for patterned tapping, where these refer to the time between taps *within* a group (within-group interval, WGI); the time *between* groups (between-group interval, BGI) was always exactly twice this (thus, the BGI was 500 ms in the 250 ms condition, and 800 ms in the 400 ms condition). These two conditions will be referred to as 250p and 400p, where "p" stands for p(atterned). The full pattern (with gaps) takes 8 beats; this pattern was tapped for 3 full synchronization patterns, and then the trial continued (without pacing tones) for a time equivalent to 7 full patterns. A 250p trial took 21 seconds, and the 400p trial took 33 seconds. Importantly, trials were matched across rates by number of tapped patterns, not total elapsed time.

There were six feedback conditions: synchronous ("S"), silent ("N": masking noise only, no sound for subject's taps), and delays of 100, 200, and 300 ms. An sixth exploratory condition ("R" for random) was run, in which the feedback delay was unpredictable (for each keystroke, a delay was randomly chosen from 100, 200, or 300 for that keystroke). The data for this sixth condition will be presented in the graphs, but will not be analyzed further.

All the trials for a particular rate and pattern were run together as a block. Within each block, the 18 trials were randomized for each subject. The total times for the 18 trials for each block type were as follows: 250p: 6.3 minutes; 250r: 4.5 minutes; 400p: 9.9 minutes; and 400r: 6.9 minutes, for a total tapping time (excluding practice) of 27.6 minutes. The two blocks for a given tempo were always placed tegether, although half the subjects received the 400 ms rate first and half the 250 ms rate. Additional counterbalancing had the patterned blocks preceded the regular blocks for half the subjects, and half vice-versa, although for any given subject both patterned blocks always either preceded or followed the regular blocks.

4.1.5 Analyses

This section describes some general aspects of the data analysis. More precise discussion of data filtering and extraction of measurements is provided in Appendix B.

Data Filtering

Subjects were replaced if they consistently showed inadequate performance on the synchronization phase, or when they reported themselves as falling asleep and this was evidenced as gaps in the data; this occurred rarely. Individual trials in which adequate synchronization was not achieved were excluded from further analysis; in no case did this filtering result in an empty cell for a subject in a condition. Only the data in the continuation phase of a trial was analyzed, and exceedingly short ITIs which resulted from the brush of an adjacent key were removed. Other than this, no outliers were filtered.

Post-Hoc Tests

For post-hoc pairwise comparisons, a procedure appropriate for testing all pairs of means was used, specifically, the Newman-Keuls test for pairwise significance at the .05 level. All the reported pairwise tests in this dissertation will use this procedure.

Grouping

The raw data files from the experiments contain full timing information for all of the subject's keystrokes, as well as timing information for both the pacing and the feedback tones. A necessary precursor to almost any analysis of the patterned tapping data (be it errors or timing) is delineation of the groups of the tapped pattern in the experimental data. A 4-2 pattern is characterized as two groups of different lengths (a 4 group and a 2 group), separated by a longer time interval. Extraction of groups from the tapping data was implemented in the following manner (using the 250p condition as an example). Recall that the nominal WGI in this condition(i.e., that given by the pacing tone during the synchronization phase) was 250 ms, with a BGI of 500 ms. An ITI cutoff was established halfway between these two values (i.e, for the 250p condition at 375 ms, and similarly for other rates); ITIs greater than 375 were classified as BGIs, and ITIs less than 375 were classified as WGIs. Thus, any taps separated by more than 375 ms would be classified as belonging to separate "groups". There is some extent to which such a classification is arbitrary (why 375 ms instead of 350 or 400?), but it also seems fairly logical. In general, such automated grouping measures closely matched measures derived from visual inspection.

4.2 Results: Patterned Tapping

The conditions of interest are the 0 ms (SAF), N, 100, 300, and 300 conditions, although a sixth exploratory condition ((R)andomized delays) was also run. The data from this condition is presented for informational purposes in all the data graphs, but for expository purposes, the statistical analyses will only include the main 5 conditions. The exclusion does not affect which differences are significant.

As a more detailed example of the experiment, Figure 4.1 provides an example of all three trials for 1 subject in the 250 ms DAF condition. This is a moderately impaired subject, in what will be seen to be a fairly impairing condition.

4.2.1 Counting Errors

One particular impairment reported in previous tapping studies (see Section 2.4) is counting or number errors: the extent to which groups are performed with an

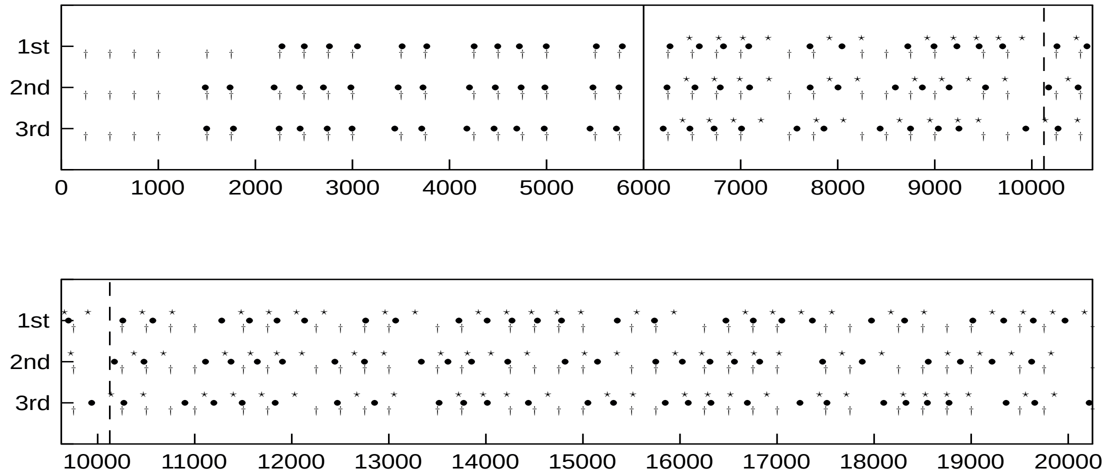


Figure 4.1: Experiment 1: Sample trials, 250p condition, 200 ms DAF. The x-axis is milliseconds from trial start; for clarity, the plots for each trial extend across the two boxes. The solid vertical line in the top part delineates the synchronization and continuation parts of the trial. The vertical hyphenated line indicates the area of overlap between the two sections. Dots indicate subjects taps. Daggers in the synchronization phase indicate pacing tones; daggers are also indicated in the continuation part for reference, but do not indicate sounded tones. Feedback tones in the continuation phase are indicated by asterisks. Some obvious errors are circled.

incorrect number of taps. Such measures are affected by DAF. Note that a given group may be mistapped because one or more taps were *omitted* (deleted), or because one or more taps were *inserted*. For a 2 part pattern (e.g., the 4-2 pattern used here), these can be separately tallied for each of the groups in the pattern.

A number of error measures for the patterned conditions of Experiment 1 are provided in Figure 4.2.[3] These measures are defined as follows (see also Appendix

[3]Most graphs contain bars for standard errors, although these were sometimes omitted when it made the graph too cluttered. These error bars are between-subject error bars, even though the actual statistical analyses are within subjects.

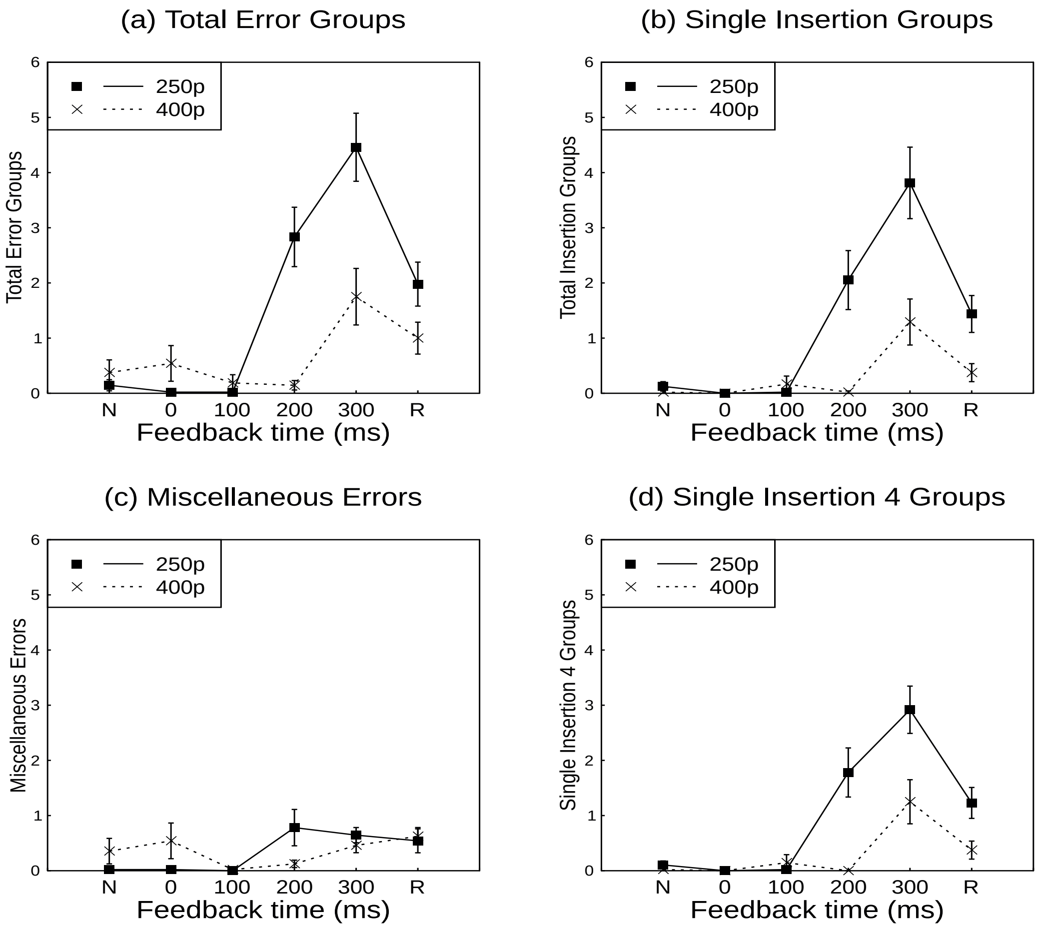


Figure 4.2: Experiment 1: Patterned Tapping, Error Measures

B for some further details):

1. Total Errors (Figure 4.2.a): Any group which did not include either 4 or 2 taps was counted as one error (thus, all deletion groups and all insertion groups). Although one could do more sophisticated analyses keeping track of the ordering of groups, this turns out to make almost no difference.

2. Total (Single) Insertions (Figure 4.2.b): This measure tallies the total number

of groups which contains a single extra tap. Here, this includes any group of 5 taps, and any group of 3 taps which follows a group of 4 or more taps. Note that a 3 group is potentially ambiguous between a 4 group with a deletion, and a 2 group with an insertion; the check for a long group is a way of resolving this ambiguity, although one which potentially *underestimates* the number of insertions on the 2 tap portion of the pattern (again, see Appendix B.

3. Miscellaneous Errors: (Figure 4.2.c): *All* counting errors except for single insertions. Thus, this includes single and multiple deletions, and multiple insertions. Note that (by definition) the Total Errors are the sum of Single Insertions and Miscellaneous Errors.

4. 4-group Insertions: (Figure 4.2.d): Total of all groups of 5 taps (that is, the Total Insertions Measure without the 2 groups).

All these measures are displayed as (mean) total count within a trial. Since trials at different rates were equated for number of *patterns*, not total elapsed time, this allows for direct comparisons across different rates within an experiment, since the number of possibilities for error is identical. It also allows for comparison across experiments in those cases where the number of patterns is the same (which is the case for Experiments 2, 3, and 4; Experiment 1 is slightly shorter).

Inspection of Figure 4.2 shows that single insertions are the predominant error (with very few miscellaneous errors). In addition, the majority of the insertions occur in the 4 group. Overall, the 400p condition appears less impaired than the 250p condition.

These conclusions are verified by statistical analyses. For each error measure, a single-factor ANOVA was performed with feedback condition as the independent variable; when significant, this was followed by a full set of pairwise comparisons using the Newman-Keuls procedure with a significance level of .05.

For the 250p condition, the omnibus ANOVA was highly significant for the overall error measure ($F(4, 60) = 34.15$, $p < .001$) and for the total insertions measure ($F(4, 60) = 21.71$, $p < .001$). For both measures, post-hoc tests showed no differences between 100, 0, and N, which all were significantly less than both 200 and 300; 200 was also less than 300. Of the insertions, 4 group insertions (overall mean, .96) occurred more often than unambiguous 2 group insertions (overall mean, .24), out of a total insertion count of 1.20 (see Appendix B for data showing that this is not an artifact of the counting algorithm). Total miscellaneous errors (everything except single insertions) had an overall mean of .29 (out of a mean for total error of 1.50). Although the miscellaneous errors are small in magnitude, they were significantly affected by the delay manipulation; an omnibus ANOVA gave ($F(4, 60) = 6.22$, $p < .001$), with posthoc tests showing that 100, N, and 0 were significantly less than 200 and 300.

In the 400p condition, the omnibus ANOVA was significant for the overall error measure ($F(4, 60) = 6.88$, $p < .001$) and for the total insertions measure ($F(4, 60) = 8.57$, $p < .001$). For both measures, post-hoc tests showed that the 300 condition had more errors than all other conditions (which showed no significant differences). Of the insertions, 4 group insertions (overall mean, .28) occurred more often than unambiguous 2 group insertions (overall mean, .02). Total miscellaneous errors (everything except single insertions) had an overall mean of .30 (out of an overall mean for total error of .60)

As can be seen, insertions are the predominant error. This is true in spite of the fact that deletions are probably being overestimated, since reported deletions may occur as a result of a slowdown which (based on visual inspection) might show clear grouping, but which exceeds the BGI cutoff established for automatically delineating groups. Perhaps more importantly, deletions will be reported when a subject's keypress is insufficient to trigger the key. [4]

Combining the data for the two tapping conditions (250p and 400p), and performing an ANOVA with tapping condition as a factors showed that for total errors, single insertions, and single insertion 4 groups, the measures in the 250p condition were significantly higher than in the 400p condition (all p values $< .005$). Miscellaneous errors did not differ significantly. However, any attempt to conclude that the 250p condition is more difficult than the 400p condition is weakened by a possible range problem: it may be that the delays used in this experiment do not cause maximal impairment. This will be addressed in later experiments.

4.2.2 Timing Measures

Various timing measures have been reported in the literature as being greater (i.e., slower) under DAF than under SAF, for both speech and tapping. Two timing measures will be reported for the Experiment 1 data: an overall mean of all the WGIs (regardless of what sort of group they occurred in), and an overall BGI mean; these data are shown in Figure 4.3.

In the 250p condition, the omnibus ANOVA was significant for the overall WGI measure ($F(4, 60) = 15.77$, $p < .001$) and for the overall BGI measure ($F(4, 60) = 14.18$, $p < .001$). Post-hoc tests showed no differences between S and N (both near 250), while these were both faster than all 3 delay conditions (all above 260 ms, and not significantly different from each other). For the BGI measures, 0, N, and 100 (all near 500 ms) were all significantly less than the 200 and 300 delay conditions

[4]This was observed to happen on a number of occasions, since keyboard registration on the Fatar requires that the key cross the the halfway point on the downstroke, which necessarily requires that it also return to near the top on release. When subjects were observed to be making insufficient key depressions, they were corrected but the data was not discarded.

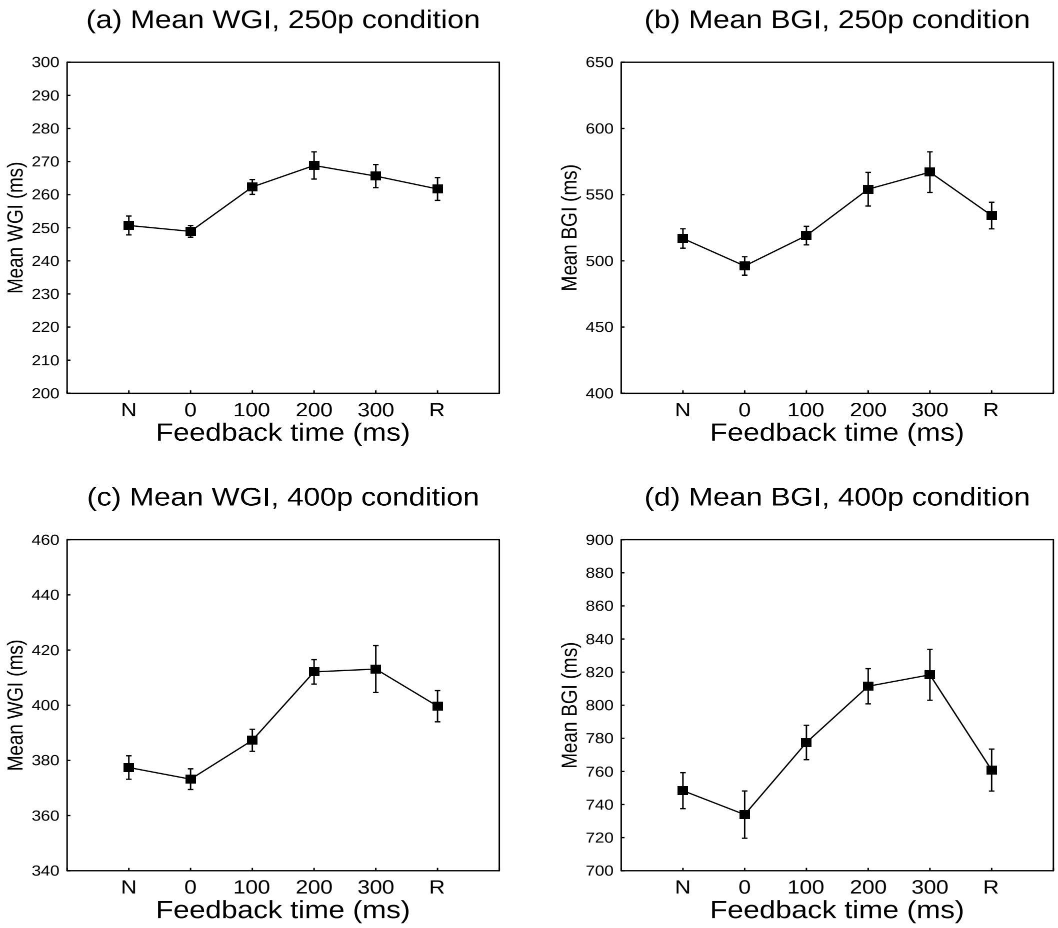


Figure 4.3: Experiment 1: Patterned Tapping, Timing Measures

(both above 550).

In the 400p condition, the omnibus ANOVA was highly significant for the overall WGI measure ($F\ (4,\ 60) = 21.77$, $p < .001$) and for the overall BGI measure ($F\ (4,\ 60) = 16.67$, $p < .001$). Post-hoc tests showed that 0, N, and 100 were all less than 200 and 300 (S was also significantly less than 100). For BGI, post-hoc tests showed that 0, N, and 100 were all less than 200 and 300 (both 0 and N were also significantly less than 100).

It can be seen that the time differences in all cases are relatively small, e.g., much less than the 33% slowdown reported by Chase et al. (1961a). In addition, this time measure (which is *not* directly confounded with insertions, unlike measures of total pattern time) nonetheless tends to *increase* in the delay conditions (which, as Figure 4.2 shows, are also where insertion errors increase). Thus, at this level errors and rate positively covary, and there is not a confound involving a speed-accuracy tradeoff.

However, as discussed in Appendix B, there *is* a confound in interpreting the rate for the 0 condition, as it may be subject to a speedup artifact. Therefore, rate measures will not be used in later analyses. This analysis, however, has been important in demonstrating that the timing does not compromise the error measure.

4.2.3 Keystroke Measures

Two other measures of interest from the literature are key velocity (roughly key force), and key-down time (dwell); these have both been reported as affected by DAF in tapping, and dwell might be roughly taken as equivalent to increased phonation time in speech. Figure 4.4 provides these values for both conditions.

An ANOVA on the mean velocity measure did not show significance for either the 250p condition ($p = .12$) or for the 400p condition ($p = .08$).

An ANOVA on key down time (dwell) was significant in the 250p condition ($F(4, 60) = 5.72$, $p < .005$), but not in the 400p condition ($p = .41$). For the 250p condition, all of 0, 100, 200, and 300 were less than the N condition.

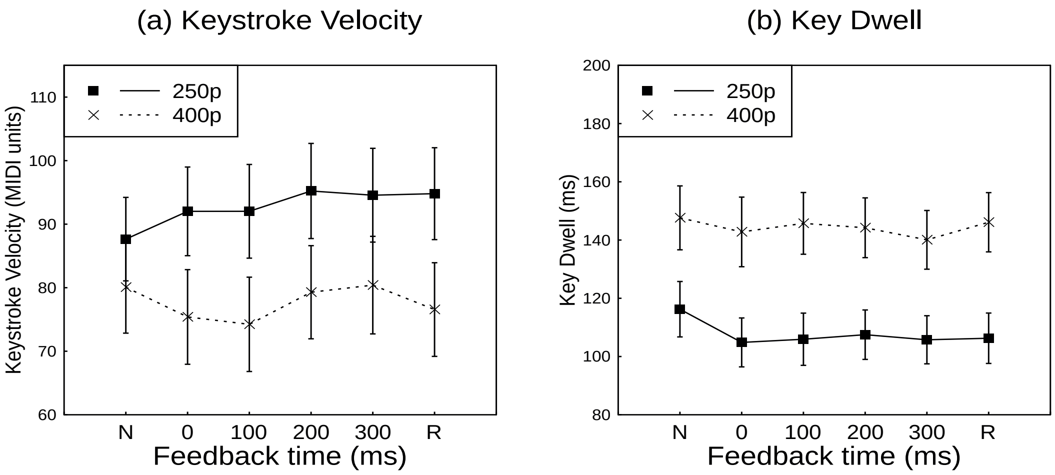


Figure 4.4: Experiment 1: Patterned Tapping, Keystroke Measures

4.2.4 Summary: Patterned Tapping

The process of correctly counting groups of taps was reliably affected by DAF, with the predominant error being single insertions in the longer (and initial) 4 group; this was true across both tapping rates. However, there were fewer errors at the slower tapping rate. For both the 250p and 400p conditions, insertions in the 4 group are more common than in the 2 group, even allowing for the underestimation inherent in the 2 group insertion count.

Feedback absence, however, did not affect counting at either rate (relative to SAF), nor did the 100 ms delay condition.

It was noted earlier that number errors and higher level timing errors have sometimes been confounded, since insertions can lead to longer group times. The current data suggests that this confound is not significant, and that timing and number measures positively covary. For the 250p condition, counting errors and slowdowns were *both* highest at the 200 and 300 ms delays; for 400p, they were both highest at

300 ms.

Timing, velocity, and dwell will not be used as measures in later experiments, for reasons discussed above. The error data is quite robust, however, and detailed timing measures will occasionally be introduced later when they address a particular hypothesis.

4.3 Results: Regular Tapping

For isochronous tapping, it is not possible to look at counting or grouping errors errors (since there are no "groups"). However, it is possible to apply additional timing measures. Again, the random ("R") condition is provided for informational purposes in the graph, but is excluded from further analyses.

4.3.1 Timing Measures

For regular tapping, there is only one type of intertap interval (ITI) to be analyzed (unlike the patterned case, where there are separate BGIs and WGIs). With only one interval, it was also worth looking at the *variance* of ITIs within trials; e.g., DAF might lead to wider *instability* in tapping regularity even if the mean was not affected. As a measure of this, for each trial the (sample) standard deviation of the ITIs for that trial was calculated, and this was used as a data point in an ANOVA. Figure 4.5 shows the mean ITI values for the regular tapping conditions, as well as the standard deviation for each.

The ANOVA for mean ITI was significant for both the 250r and 400r conditions. For the 250r condition, the omnibus ANOVA gave $F\ (4,\ 60) = 7.91,\ p < .001$. Post-hoc tests showed that the mean ITI in the 200, 0, and N conditions (around 248 ms) was slower than in the 300 and 100 conditions. For the 250r condition, the variance measure did not significantly differ ($F < 1.0$).

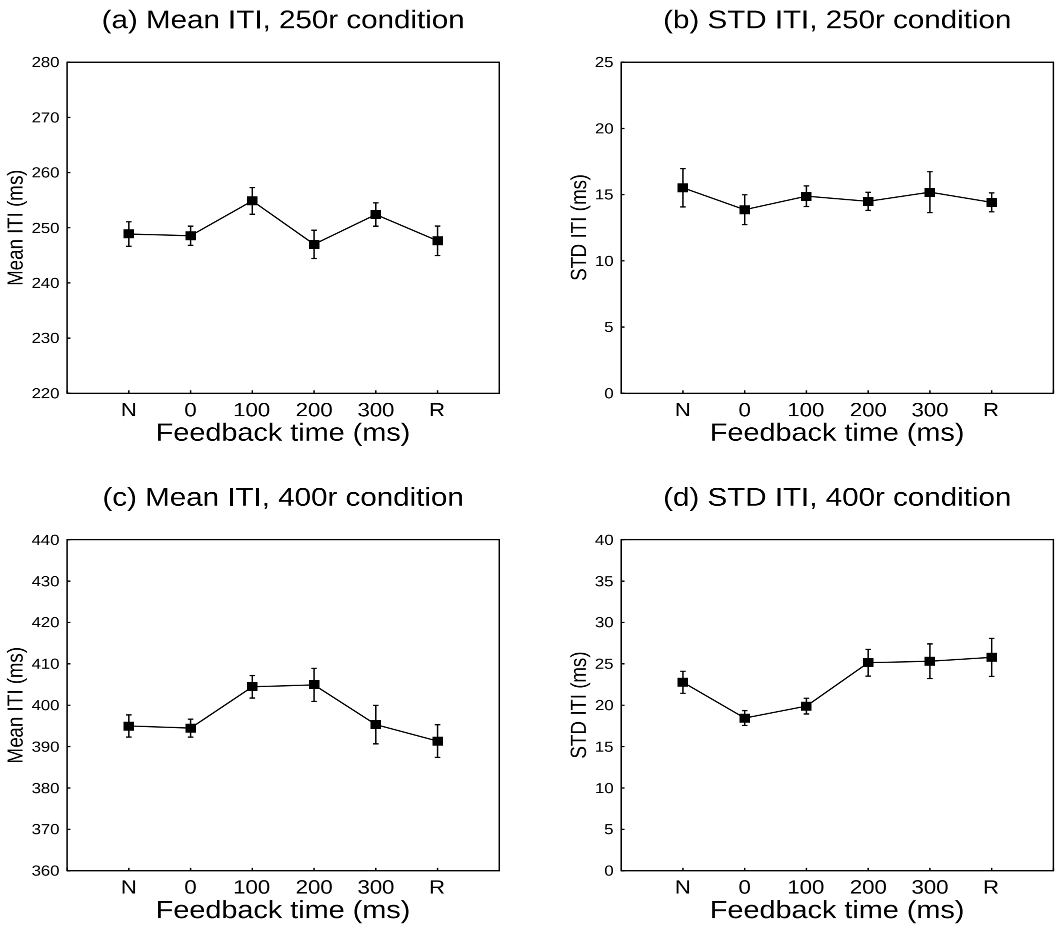


Figure 4.5: Experiment 1: Regular Tapping, Timing Measures

For the 400r condition, the omnibus ANOVA gave F (4, 60) = 3.91, $p < .01$. N (395 ms) was less than the 100 ms delay (404 ms), and 300 was less than both 100 and 200. For the 400r case, the variance measure *was* significant F (4, 60) = 4.96, $p < .005$). Both 0 and 100 were less than both 200 and 300.

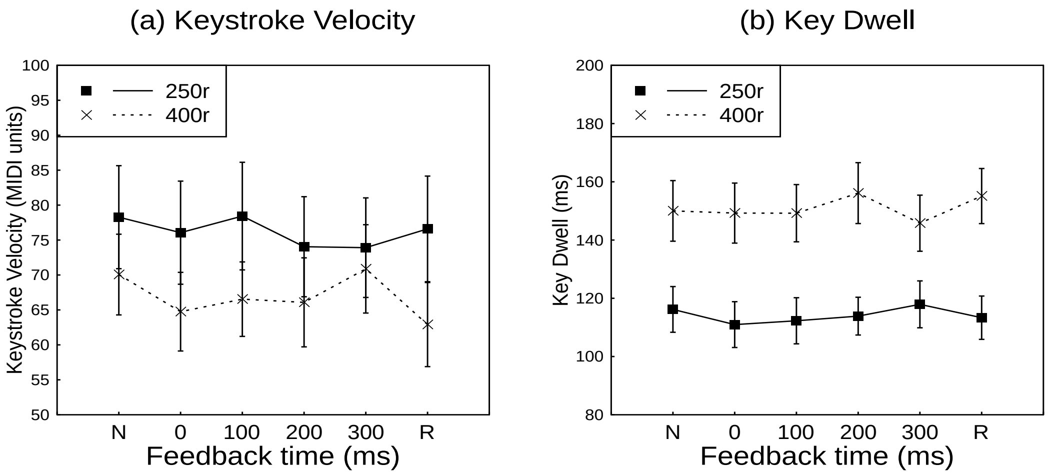


Figure 4.6: Experiment 1: Regular Tapping, Keystroke Measures

4.3.2 Keystroke Measures

Velocity and dwell measures are also straightforwardly applied to isochronous tapping. Figure 4.6 provides these values for both measures and both conditions.

For the 250r condition, an ANOVA did not show significant differences for either velocity or dwell, although both could be viewed as marginal (the p value for dwell was .051, and for velocity was .09). Post-hoc comparisons showed no pairwise differences. For the 400r condition, an ANOVA did not show significance for either measure ($p > .20$).

4.3.3 Additional Timing Measures

Additional timing measures were applied to the regular tapping data. One such measure was the extent to which subjects slowed down over the course of a trial. This tempo change measure was implemented as the difference in mean ITI between the first 5 continuation taps and the last 5 continuation taps. In the 250r condition, the changes were small in all conditions (less than 5 ms), and did not differ between conditions. In the 400p condition, the value for 300 ms delay was larger (a slowdown of 19 ms), and this value was significantly larger than all other changes (which were under 5 ms, and did not significantly differ).

The second measure is autocorrelation (correlation between different values from the same time series), which has been used in many studies of isochronous tapping. For example, Wing and Kristofferson (1973) use it to separate out motor and central components of timing variance. This will be applied in an exploratory fashion, as it is not obvious what the predictions would be in the DAF case.

Autocorrelations in the 250r case are all rather low (all the values at lags 1, 2, and 3 are less than .15). In the 400r case, the autocorrelations are also all less than .15, except the Lag 1 autocorrelation in the 0 condition, which is -.25 (consistent with the presence of a negative correlation predicted by the Wing and Kristofferson model). This differs significantly from all the other values, providing tentative evidence that DAF (and feedback absence?) can interfere with this measurement.

Note that in the literature, auto correlation measures are often applied to highly practiced subjects, in which the base variance is already low; this may contribute to the absence of a strong autocorrelation measure. Also, Russell Church (personal communication) points out that autocorrelation measures are affected by drift; although the speed changes here are small, this may be relevant.

4.3.4 Summary: Regular Tapping

The analyses of the regular tapping data are fairly weak. The 250r condition showed a slowdown at 100 ms, relative to 0; 100 ms (and also 200) were also slow for 400r. The only condition showing a statistically significant slowdown over time was the 300 ms delay in the 400r condition. There are no consistent signs of a peak for 200 or 300 ms, contrary to the patterned cases.

4.4 Summary and Discussion

4.4.1 Replication

One issue for the current experiment was to determine the extent to which use of the continuation paradigm confirmed earlier results in the literature. Some robust confirmations of prior reported results were the following:

1. There is severe impairment in patterned tapping under DAF.

2. There is less impairment in regular tapping than in patterned tapping, although this may be because the most useful measure (insertions) doesn't apply here. Alternatively, perhaps regular tapping is under a different (automatic) form of control than patterned tapping.

3. Auditory feedback absence did not show any consistent impairment effects, in either regular or patterned tapping. For error counts in patterned tapping, there were no significant differences between feedback absence and synchronous feedback. Of the 4 available timing comparisons, only one (BGI in 250p) showed a difference for these two conditions.

4. Insertions were the quantitatively predominant error for patterned tapping; this has been reported only in an anecdotal fashion in earlier studies).

Some important differences were the following:

1. Timing measurements were not generally significant here, unlike prior studies. This may be due to the more precise timing resulting from the continuation paradigm. However, crucially, there is no evidence that timing and errors are

confounded (i.e., trade off). Thus, errors can be looked at without regard to timing.

2. Key velocity (force) measurements were not signficant; this interesting difference will be discussed below.

4.4.2 New Results

Although the delays used did not provide enough range to locate a critical region at either tapping rate, there are some other new results:

1. Insertions are maximal on the 4 group in the 4-2 pattern (compared to the 2 group).

2. Insertions are the predominant error; furthermore, the dominant type of insertion involves a *single* additional tap (multiple taps are included under Miscellaneous errors, which are low).

3. Typical DAF impairment occurs even in the absence of SAF, suggesting that the bone-conducted feedback in speech may not be central to the impairment (i.e, that the impairment is not due to conflicts within the auditory feedback channel).

4.4.3 Velocity/Force Measurements

Keystroke velocity was not reliably affected by DAF in the current experiment, nor is it consistently significant for other experiments reported later in this dissertation.[5]

---

[5]Note that key force (or loudness) is not simply an arbitrary measurement, but plays a significant role in some theories of DAF. For example, one proposal is that increased force is a way of increasing feedback in the undisturbed (kinesthetic, tactile) channels. However, an alternate explanation for the force increase is that it is

This is a bit of a surprise, given the robust reports of loudness increases in speech under DAF, and of increase in key force for tapping under DAF. Inspection of Figure 4.6 does seem to show a trend for higher velocity in the 200 and 300 ms delay conditions, but it is not reliable due to the high variance.

Why is there this difference in results from other studies of DAF and tapping? There are a number of possibilities.

1. It may be due to the use of a *velocity* measure (provided in the MIDI signal) and its particular implementation on the Fatar keyboard, contrasted with the more precise force measurements used in, e.g., Chase et al. (1961b).

2. Earlier studies have often used spring-loaded keys (offering some resistance), in contrast to the relatively low resistance key here. Lack of resistance might increase variance.

3. Sound level did *not* vary with keystroke velocity here, and this may be a factor, although other studies have reported force increases in such circumstances (e.g., Ruhm and Cooper, 1963).

4. Force measurements in earlier studies have often used selective measures, such as the keystroke of *maximal* force within a group, rather than taking the mean across all keystrokes, as was done here.

In any case, key force will not be reported as a measure in later experiments.

4.4.4 Implications

This experiment has established some bases for the later experiments:

---

simply a stress response. This might still be of interest but is less central for theories of the DAF phenomenon.

1. Patterned tapping should be used, as it is robustly affected by DAF. Use patterns.

2. Of the measures available, errors are the most robust at showing differences, and of the errors, insertions are the most robust.

3. Errors can be looked at without regard to timing (which showed only small differences, and positively covaried with errors).

Chapter 5

Experiments 2 and 3: Critical Interval

This chapter will report on two experiments addressing two related questions raised earlier: (1) is there a critical interval (delay of maximal impairment) in tapping, and if so, (2) does its value shift with tapping rate? The first question is clearly a preliminary to the second, and also bears on the issue of the extent to which speech and tapping demonstrate similar impairment: do both domains show a critical interval? The second question is relevant to two particular theories. First, as described earlier, MacKay's Node Structure Theory postulates nodes with a fixed, non-shifting refractory period, and predicts that critical interval will not shift with tapping rate. Second, feedback theories which hypothesize that an act is repeated if confirming feedback is not received with a fixed period of time would also predict there would be no shift. Although Experiment 1 showed a clear increase in impairment with delays up to about 300 ms, it only tested a limited range of delays, and thus may not have reached the delay values at which performance might decline.

5.1 Experiment 2

Experiment 2 used only a 4-2 patterned condition and investigated the question of the relation between tapping rate and amount of delay. Except for differences noted here, the procedure and methods were identical to Experiment 1.

5.1.1 Method

Equipment

The synchronization tones were hard to hear in the practice trials which used SAF, so the loudness for the synchronization (pacing) tone was raised from 79 to 80 dB.

Subjects

12 subjects participated in this experiment.

Conditions/Stimuli

The 4-2 pattern was tapped at 3 different metronome speeds: 250 ms, 350 ms, and 450 ms (the 250 ms serves as a partial replication of Experiment 1). The delay conditions were 100, 200, 300, 400, and 500 ms fixed delay, synchronous feedback (0 ms delay), and one additional exploratory condition which will be omitted from any further consideration. Thus, there were 3 blocks, with 21 trials in each block. Each of the 6 possible block orders was presented to two subjects.

Due to task difficulty at the faster rates, an additional synchronization pattern was added to create 4 pacing patterns for the 250p and 350p conditions; 3 pacing patterns were used for 450p. The continuation portion was also extended by a few beats (to increase the chance of getting a full pattern if subjects' tapping rate slowed down), so the revised continuation period allowed for 7 1/2 full patterns rather than the 7 of Experiment 1. This invalidates direct (unnormalized) comparisons with Experiment 1; however, this longer length was maintained for all remaining experiments.

Total times for individual trials were as follows: 250p: 23 seconds, 350p: 33 seconds, and 450p: 38 seconds. With 21 trials per tapping condition, total tapping times (not including breaks) were: 250p: 8 minutes, 350p: 11.5 minutes, 450p: 13.3 minutes, giving a total of 32.8 minutes of tapping, not including practice.

Some other differences from Experiment 1 were that *all* the trials were patterned (i.e., there was no alternation with regular tapping, possibly affecting how subjects might perform over time, or the overall difficulty of the experiment), and that a larger percentage of the trials involved a significant amount of delay.

5.1.2 Results: Experiment 2

Experiment 1 established insertions as a useful and representative error measure, and insertions are the only measure that will be reported for Experiment 2. Experiment 3 (addressing the same issue) will be analyzed in more detail.

Counting Errors

The insertion data for all 3 conditions are graphed in Figure 5.1.

In the 250p condition, the omnibus ANOVA was significant for the insertions measure ($F(5, 55) = 10.39$, $p < .001$). Post-hoc tests showed 0, 100, 400, and 500 to all be significantly lower than 200 and 300. Thus, at the 250 ms tapping rate the peak is near 200 or 300 ms.

In the 350p condition, the omnibus ANOVA was significant for the total insertions measure ($F(5, 55) = 9.93$, $p < .001$). Post-hoc tests showed 0, 100, 200, and 500 to be less than 400, with 0, 100, and 200 also less than 300. Thus, at the 350 ms tapping rate the peak is near 300 or 400 ms.

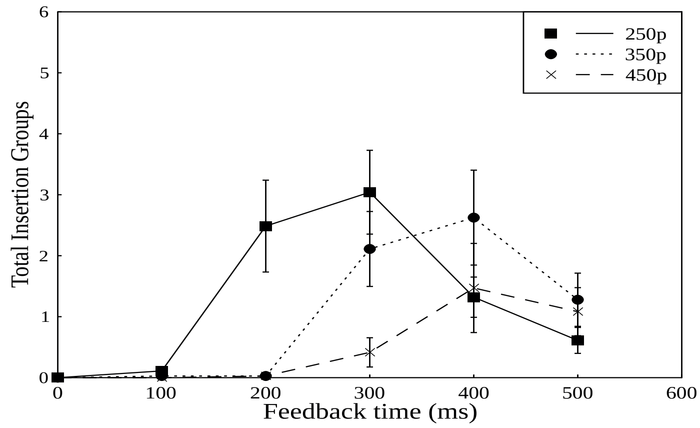


Figure 5.1: Experiment 2: Single insertion errors

In the 450p condition, the omnibus ANOVA was significant for the total insertions measure (F (5, 55) = 4.17, $p < .005$). Post-hoc tests showed no differences between 100 and 200 to be less than both 500 and 400, and 0 and 300 were less than 400. At this 450 ms tapping rate, the maximal impairment was near 400 or 500 ms, but there are not enough delays to determine if there is a decrease in impairment at higher delays.

Inspection suggests a fairly clear change of peak (based on the maximum within each condition), and a quadratic curve fit for the 250p and 350p conditions confirms this shift. Taking all all points into account, the maximum of the best-fitting quadratic occurs at 278 ms for the 250p and 520 ms for 350p. Although the 450p condition has a clear maximum, the best-fitting quadratic for the 450p condition actually gives a minimum, due to the flat portion of the curve. If the flat part of the curve is omitted (i.e., omitting the 0 point for the 250p condition, and the 0 and 100 points for the 350p and 450p conditions), the peaks of the best fitting curves are then 250p: 298 ms, 350p: 374 ms, and 450p: 458 ms, more closely matching the visual peak.

Another statistical test relevant to a critical interval shift is to test the interaction with a 2-way ANOVA with tapping rate and delay as factors. Such an ANOVA shows (as could be predicted from the graphs) main effects of both tapping condition and feedback condition (both with p values near .001). More importantly, the interaction was also significant (F (10, 110) = 8.43, $p < .001$).

The interaction, however, doesn't describe the precise nature of the relationship, merely that the effect of delay differs with tapping rate. Thus, the statistical question still remains: is the peak significantly different in the two cases? Simply showing different curve fits does not provide a confidence level. There does not appear to be a standard statistical test for this question, so the following approach was used. Each subject's delay of maximal impairment was determined for each of the three conditions of the experiment; this was entered as a data point in a one-way ANOVA using tapping condition (250p, 350p, and 450p) as a factor. When different delays produced the same maximum value, the *mean* of these delays was taken.[1] The omnibus ANOVA was significant (F (2, 22) = 4.46, $p < .05$), with the mean values being 250p: 292 ms, 350p: 375 ms, and 450p: 381 ms.

Thus, the shift of critical interval is not only visually obvious, but statistically reliable.

Interestingly, corresponding to Fairbanks and Guttman's finding for speech, the delay of maximal impairment corresponded reasonably closely to the delay at which the maximal percent of counting errors were (single) insertions. For the 250p condition, the maximal percent of insertion occurred at delays of 300 (75%) and 200

[1] Other strategies for resolving such ties were also tried; e.g., taking the lowest or highest value. Such variations did not change the significance of the result.

(72%) (all other conditions were less than 57%). 350p was less correlated, with maximal percentages at 500 (87%) and 400 (67%), and with all others below 58%. For the 450p condition, the maximum was at 300 (56%).

5.1.3 Discussion: Experiment 2

All 3 conditions showed a delay of maximal impairment close to the tapping rate. However, Experiment 2 did not precisely specify the relationship was between critical interval and tapping rate, since delays were spaced widely at 100 ms intervals, and were not explicitly chosen to relate to the tapping intervals. This will be addressed in Experiment 3.

5.2 Experiment 3

The main goal of Experiment 3 was to look in more temporal detail at the critical interval suggested by Experiment 2. Experiment 2 provided evidence that the critical period shifts with tapping rate, but the granularity was rough. For example, does the peak fall precisely *at* the tapping rate, or a little before or after? Experiment 3 used more precise granularity to look at this question. In addition, this is a fairly precise experiment, and a number of further analyses will be done on the data.

5.2.1 Method

5.2.2 Subjects

18 subjects participated in Experiment 3.

Stimuli

Experiment 3 used the 250 ms tapping rate of Experiments 1 and 2, as well as the 400 ms rate used in Experiment 1. Only 2 conditions are strictly necessary for testing a critical interval shift, and 400 ms allowed for fine granularity with some overlap in rates between the two conditions, as well as a more manageable experiment length than a slower rate.

Delays were chosen separately for each tapping rate to provide fine mapping near that rate. At both tapping rates subjects performed under feedback conditions including synchronous feedback (which will be referred to as either SAF or 0 ms delay), and a number of delays that were near the WGI value. For the 250 ms tapping rate, the delays tested were 100, 150, 200, 225, 250, 275, 300, 350, 400, and 450 ms (a total of 11 delays, including the SAF condition); these were mostly at 50 ms increments, but with 25 ms increments near the tapping rate. For the 400 ms tapping rate, the delays were 200, 250, 300, 350, 375, 400, 425, 450, 475, 500, and 550 ms (plus the SAF condition); this gives a total of 12 conditions, compared to the 11 of the 250p condition.

Trial length was defined as in Experiment 2 (that is, a continuation phase of 7 1/2 patterns, with synchronization of 4 full patterns for the 250p condition, and 3 full patterns for the 400p condition). The total time for a 250p trial was 23 seconds; three trials were run in each of the feedback conditions, giving a total of 33 trials per subject, and a total elapsed tapping time of about 13 minutes. The 400p trials were 33.6 seconds, giving a total time for 36 trials of about 21 minutes. Half the subjects got the 250p condition first, and half the 400p.

5.2.3 Results

Analysis of the results showed no significant main effects or interactions involving block ordering with the measures of interest. Since different delay intervals were used, the 250p and 400p conditions will be shown on separate graphs, except for the crucial critical interval comparison. Although insertions have been established as the key measurment of interest, because this is a central experiment some further descriptive statistics will be provided.

Errors

Data for total errors, total single insertions, miscellaneous errors, and 4 group insertions (as described in Section 4.2.1) are provided in Figure 5.2 for the 250p condition, and in Figure. 5.3 for the 400p condition.

A one-way ANOVA was performed on both the total errors and total insertions measures for the 250p condition, with delay time as the single factor. The omnibus ANOVA was significant for both the overall error measure ($F(10, 170) = 9.08$, $p < .001$) and for the total insertions measure ($F(10, 170) = 9.11$, $p < .001$). For total errors, post-hoc tests grouped 100, 0 (= SAF), and 150 together at the low end, and 400, 275, 200, 225, 300, and 250 together at the high end. For insertion errors, 100 and 0 were lowest (although not significantly different from 150 and 450), while 225, 250, 275, and 300 were grouped together at the high end. Of the insertions, 4 group insertions (overall mean, 1.32) occurred more often than unambiguous 2 group insertions (overall mean, .58). Thus, (single) insertions are the predominant error, and such insertions occur more in the 4 group than the 2. There was little impairment at low delays, e.g., no errors at 100 ms.

As in Experiment 2, the delay of maximal impairment corresponded closely to the delay at which the maximal percent of counting errors were (single) insertions.

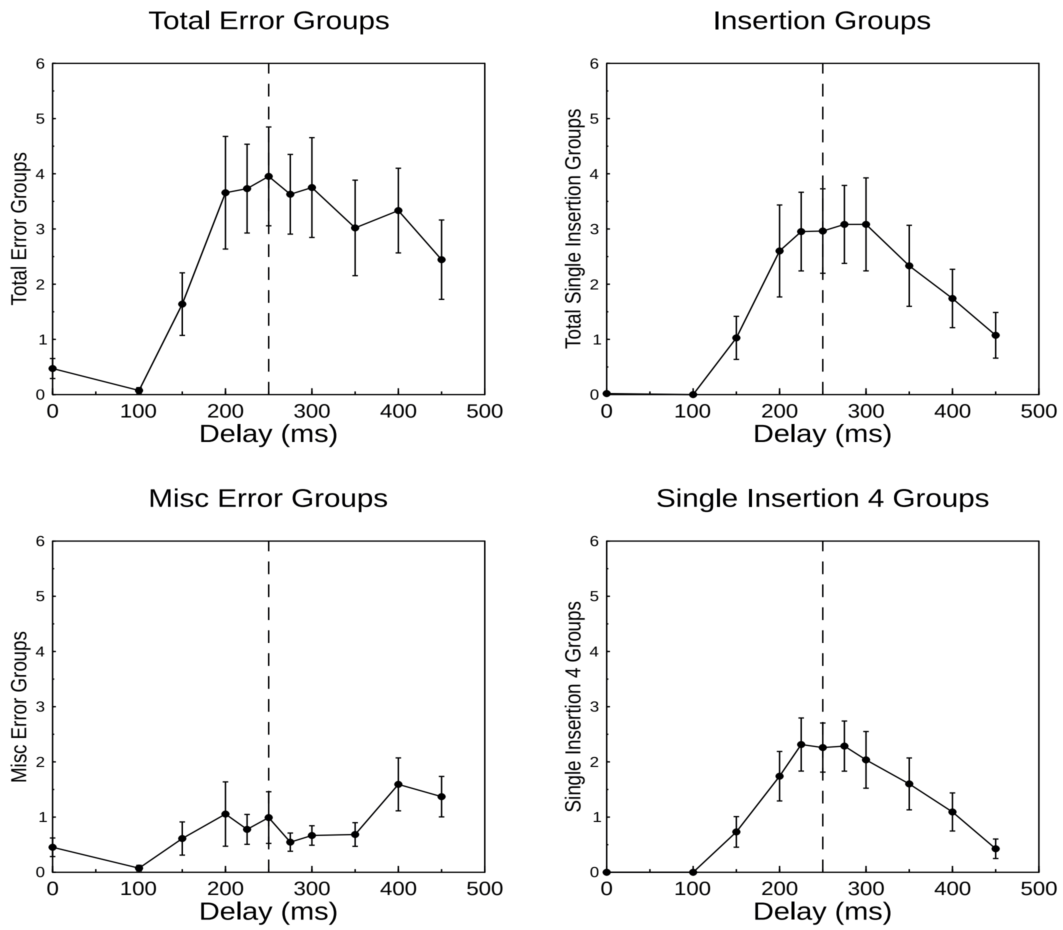


Figure 5.2: Experiment 2: Mean error data, 250p condition.

For the 250p condition, the maximal percent of insertions occurred at 275 (84%), with 225 and 250 above 80%.

The omnibus ANOVA for the 400p condition was significant for the overall error measure (F (11, 187) = 7.19, $p < .001$) and for the total insertions measure (F (11, 187) = 5.79, $p < .001$). For total errors, post-hoc tests grouped 0, 200, and 250 together at the low end; 425 was highest, with 400, 450, and 375 also high. For insertion errors, 0, 200, and 250 were lowest, while 450 was highest, followed by 400

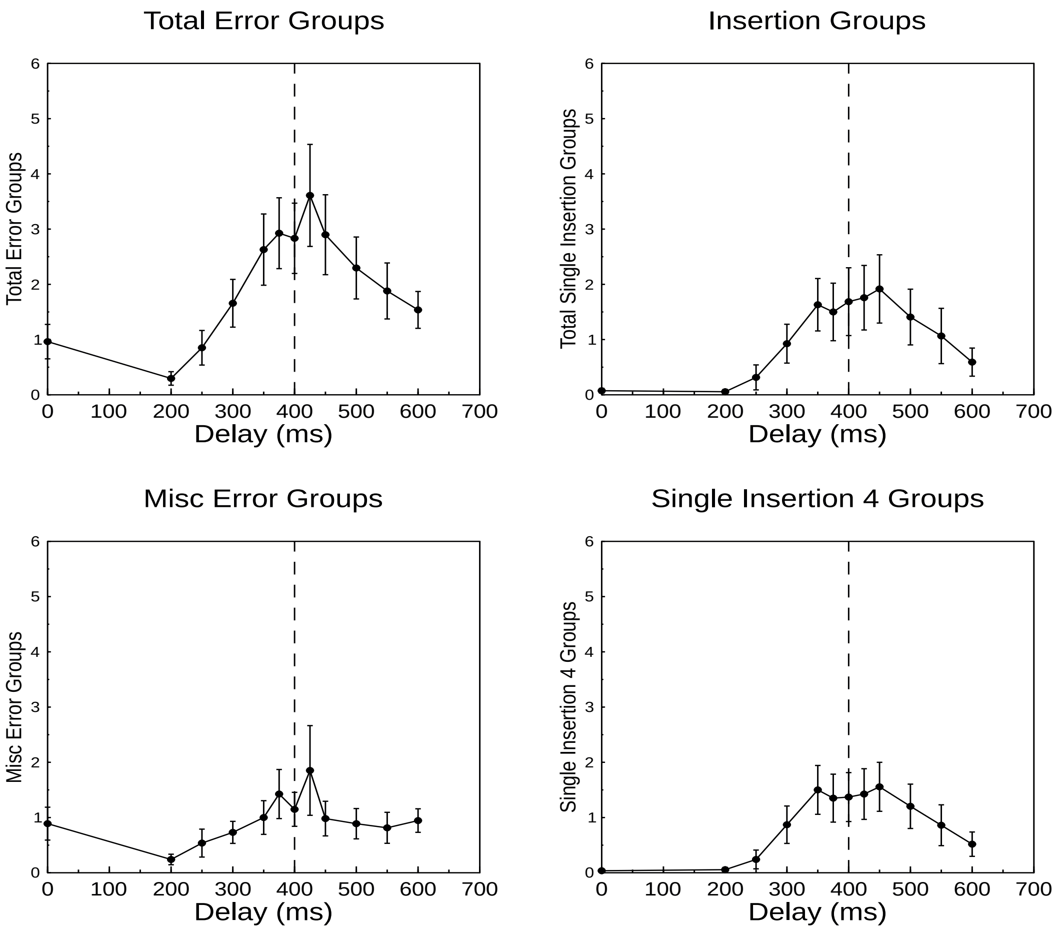


Figure 5.3: Experiment 2: Mean error data, 400p condition.

and 425. Of the insertions, 4 group insertions (overall mean, .92) occurred more often than unambiguous 2 group insertions (overall mean, .16).

For the 400p condition, the maximal percent of insertions occurred at 350, 425, and 450 (57%), with 375 and 500 above 50%.

Many results replicated across the two rates. Insertions were the predominant error measure, and both curves show a clear peak (although its location appears to be different in the two cases, as in Experiment 2). Overall, the error level in 400p

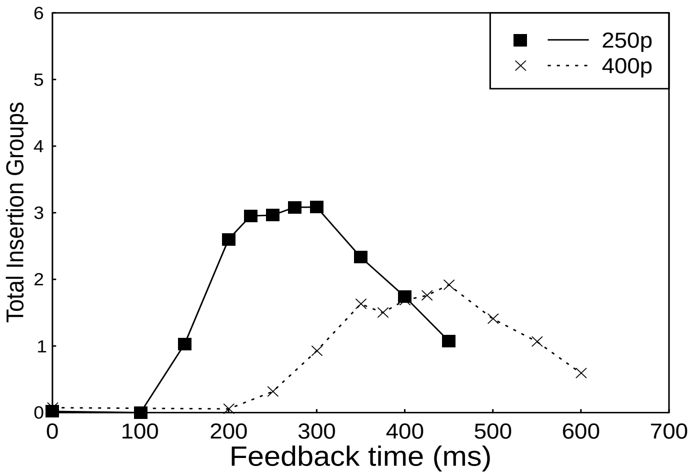


Figure 5.4: Experiment 3: Single insertion data

was lower.

Levels of impairment varied widely across subjects. In the 250p condition, subjects made virtually no insertions under SAF, but in the 250 ms delay subjects had insertion rates ranging from .33 to 12.0, with a mean of 2.96 (the value seen in Figure 5.2). Interestingly, every one of the 18 subjects nonetheless showed worse performance in the 250 delay condition on both total errors and total insertions measures (even though this was sometimes only because of 1 error in the combined three trials).

Tapping Rate and Critical Interval

Data for the insertion errors at the two tapping rates are combined in Figure 5.4.

Inspection suggests a fairly clear change of peak, and a quadratic curve fit for each of the 2 conditions confirms this. The maxima of the curves occur at 277 ms for the 250p condition and 442 ms for the 400p condition. For the 250p condition, the impairment is highest between 225 and 300 ms, extending to 200 and 350. The peak does not appear to be very precisely specified at 250, but has a range of about +/- 50 ms. If the 0 point is removed, the best fitting curve for the 250p condition has a maximum at 289 ms, and the 400p at 426 ms. The graph also shows a somewhat obvious difference with respect to where the impairment *starts*: 200 ms is highly impairing at the 250p rate, but has no effect at the 400p rate.

As with Experiment 2, a 2-way ANOVA was performed to determine whether the effects of delay vary with tapping rate. The analysis was restricted to the delay conditions shared by the two tapping conditions (S, 200, 250, 300, 350, 400, and 450). Such an analysis shows main effects of both tapping condition and feedback condition, as well as a significant interaction ($F\ (6, 102) = 7.67$, $p < .001$), showing that the effects in the two tapping rates are different.

The same statistical test as in Experiment 2 was used to verify the statistical significance of the difference in peak impairment interval. Each subject's delay of maximal impairment was determined for each of the two conditions of the experiment; this was entered as a data point in an ANOVA using tapping condition (250p vs 400p) as a factor. For the insertion measure, the omnibus ANOVA was significant ($F\ (1, 17) = 49.03$, $p < .001$), with the mean values being 271 ms for the 250p condition, and 406 ms for the 400p condition (again, fairly close to the tapping rate).[2]

---

[2]A possible concern here is that different ranges of delays were used for the two tapping conditions; this may skew the determination of the maximum impairment interval in the two different conditions. To test for this, the statistics were redone, but restricting the range for the search for per-subject maxima to the range of delays which were shared across the two conditions (200-450 ms). If anything, this strengthened the effect. The omnibus ANOVA was significant ($F\ (1, 17) = 38.00$, $p < .001$), with the mean values being 250p: 273 and 400p: 385.

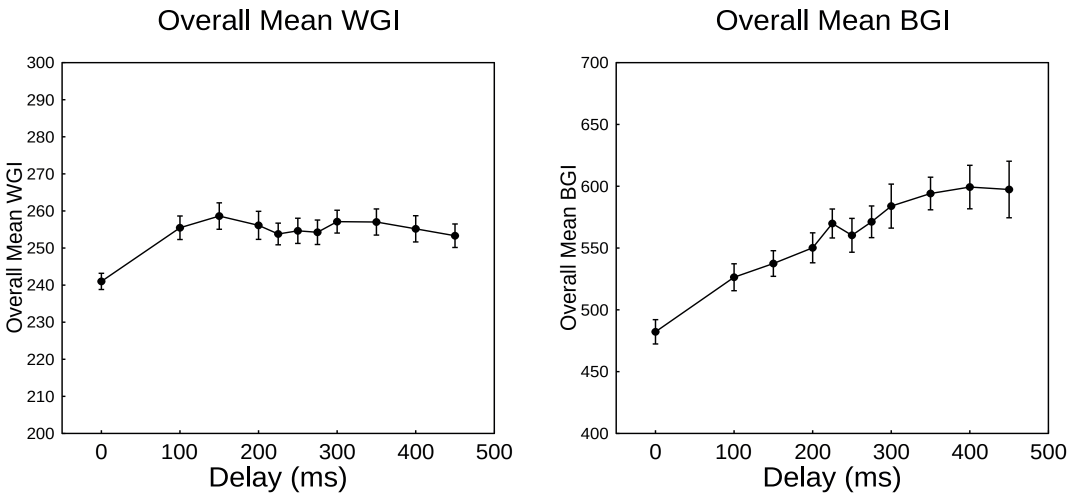


Figure 5.5: Experiment 3: Mean timing data, 250p condition.

Timing Measures: WGI and BGI

The timing data will be provided strictly for informational purposes, and to confirm that (as in Experiment 1) the differences are small, and unlikely to contribute to any important speed-accuracy tradeoffs. Graphs for overall mean BGI and mean WGI are provided in Figure 5.5 for the 250p condition, and in Figure 5.6 for the 400p condition.

Keystroke Measures

A measure that will have some relevance for later analyses of Experiment 3 is key down time (dwell). As reported in Experiment 1, this is not a robust measurement for showing differences between conditions, but the mean values will be important later. For the 250p condition, the overall mean for dwell (across all conditions) was 127 ms, ranging from 120 ms in the 0 condition to 136 ms in the 400 condition. For the 400p condition, the mean dwell of 168 m, ranging from 155 ms in the 0 condition

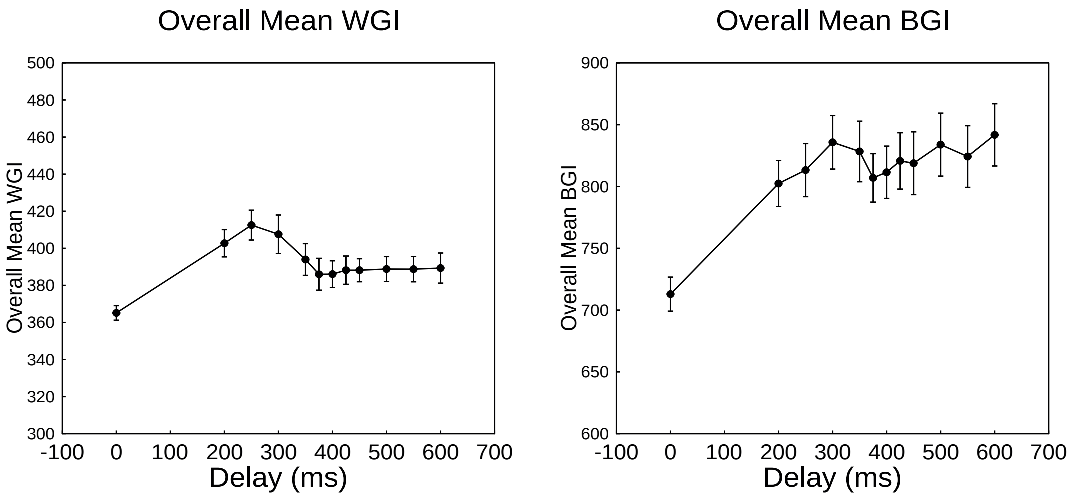


Figure 5.6: Experiment 3: Mean timing data, 400p condition.

to 174 ms in the 600 condition.

5.3 Discussion

5.3.1 Empirical Results

The results of these two experiments have established the following results:

1. The data provide the first time-controlled establishment of a critical interval in tapping. Details of timing show that the peak (at least averaged across subjects) is fairly wide.

2. This critical interval shifts with tapping rate, with a larger delay for maximal impairment at slower tapping rates (and least within the tested range of 250-500 ms or so.). Thus, there is not evidence for a fixed (magical?) delay of maximal impairment at 200 ms. Interestingly and significantly, the location of this peak corresponds closely to the tapping rate.

3. The proportion of insertion errors in tapping is highest near the critical interval, which is significant because it corresponds to a report of the same relationship for speech (Fairbanks and Guttman, 1958).

5.3.2 Theoretical Implications

The shift in critical interval for tapping provides some initial evidence against MacKay's Node Structure Theory, which suggests that the interval should be fixed. Although the theory could be modified to fit this finding (it is not obvious that the hypothesis that refractory period is fixed for a node type is central to the NST), a more telling argument against the NST will be provided in Chapter 6.

Similarly, the fact that slower rates have a longer delay of maximal impairment is an argument against any theory of feedback guidance that suggests that feedback must be received within a fixed time period for optimal performance, such as Chase et al. (1961b).

The fact the the critical interval robustly shifts in tapping also suggests that, given the already conflicting results for speech, the claim that the critical interval does not shift for speech needs to be looked at more closely.

Chapter 6

Experiment 3: Timing Details

Chapter 5 established the presence of a critical interval in tapping, as well as the shift of this interval with (and closely matching) the tapping rate, with most impairment occurring when the delay interval is close to the ITI. This poses problems for MacKay's Node Structure Theory, and for fixed-time based theories of feedback guidance, both of which predict the critical interval will stay fixed.

The current chapter will look in more detail at the precise timing relationship between feedback and insertions in Experiment 3, and their implications for both the NST and the Displaced Rhythm Hypothesis. Such timing analyses have not been done before in the DAF literature.

These detailed analyses of insertion groups will be done on the Experiment 3 data only, and will be restricted to insertions in the 4 groups (which, as the data in Experiment 1 and 3 demonstrated, are where most insertions occur). The analyses could, of course, have been done on any experiment, but there is no reason to believe the results would be any different for the other experiments.

6.1 Insertion Group Timing: Anticipation Relationships

In MacKay's Node Structure Theory, the sound serves to restimulate the node which caused that sound, leading to a repetition. One particular prediction of MacKay's NST (and, more generally, of any stimulus-driven theory of DAF insertion) is that a particular stimulus event (the feedback) will precede the inserted tap, in some appropriate timing relation. Section 3.3.3 provided evidence that for the theory to apply to speech, the repetition (insertion) must almost immediately follow the feedback onset (i.e, with a time close to 0 ms).

The tapping data can be analyzed to see if this prediction holds. Figure 6.1 plots some sample groups in which an insertion has occurred (that is, groups of 5 taps) for the 250p tapping condition of Experiment 3. Figure 6.2 includes similar data for the 400p tapping condition. These particular DAF values (250 and 400 ms, respectively) were chosen for display because they are highly impairing conditions. These should be good tests for any theory of insertions. The choice of presented plots is semi-random, being simply the first insertion group in this condition for each of the first 3 subjects.

Assume that the inserted tap is the final tap (that is, that subjects correctly tap 4 times, and then for some reason insert a fifth tap). None of these graphs provide any obvious example of this tap being preceded by close feedback. A convincing argument, of course, requires analysis of the entire data set. To address this issue precisely, one question is what amount of time should be expected to occur between feedback tone and the repetition it causes; as described above, this needs to be almost instantaneous for speech, and one would a-priori expect the same timing relation to apply for tapping. A second question is, what part of the feedback tone should trigger the tap? Again, the speech data suggests it will must be the *onset* of

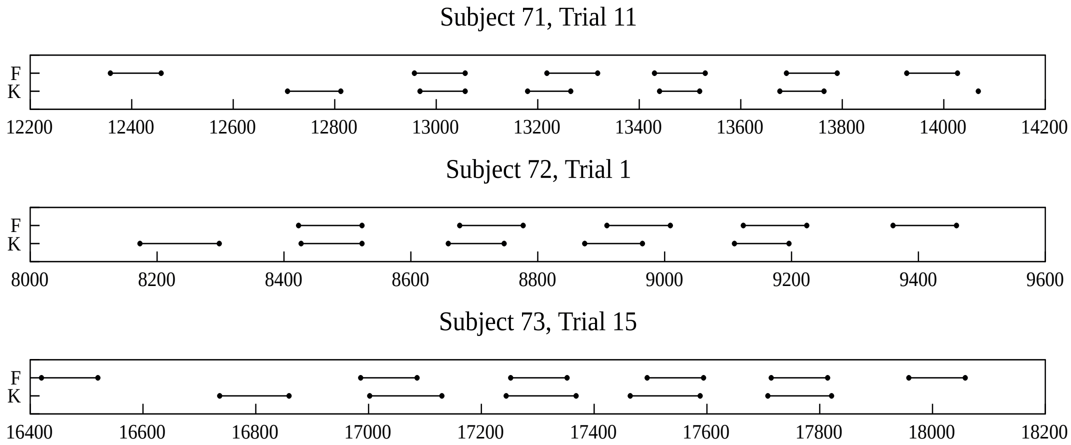


Figure 6.1: Experiment 3: Sample insertion groups, 250p condition, 250 ms DAF. The lower line (labelled 'K') indicates subjects' keystroke onsets and durations; the 'F' line indicates feedback tones (of fixed 100 ms length). It should be noted that the horizontal axes for the different trials are not identically scaled, since the goal was to place all of the relevant insertion group information into the same physical space. The X-axis denotes time from the start of the trial.

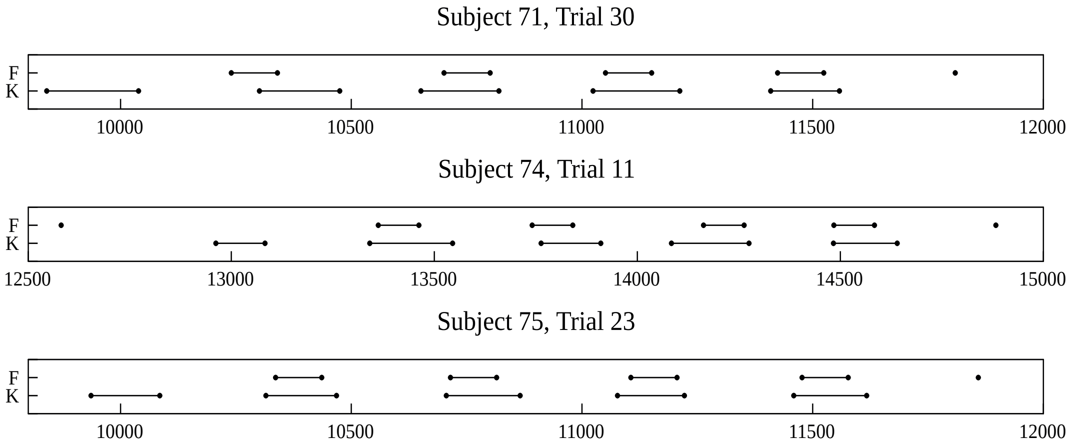


Figure 6.2: Experiment 3: Sample insertion groups, 400p condition, 400 ms DAF.

the feedback. [1]

[1] Onsets of events would also be likely to be perceptually salient. It is not plausible that the stimulus is just anything which stimulates a (possibly-relevant) node: for instance, continuous sound will stimulate the auditory channel, but seems unlikely

Continuing to assume that the final tap is the insertion, and looking at the 250p condition, it is possible to look at the distribution of timing for the closest feedback onset preceding the tap for the 250 and 275 ms delay conditions (both in the vicinity of maximal impairment). The mean time between feedback and tap at 250 ms delay is 136 ms, with 49% of the times being greater than 200 ms. For 275 ms delay, the numbers are greater: the mean is 183 ms, and 47% are greater than 200 ms. Such times are much greater than the approximately 0 ms response times shown to be necessary for speech. This analysis specifically addresses delay conditions with high impairment (and, furthermore, ones which emphasize long timing relations). It is possible that the timing differences will be less in conditions with other (and less impairing) delays. But note the logic of the argument: if there is a case where there are many insertion taps and the hypothesized timing relation doesn't hold, then the theory is unlikely to be correct, even if the hypothesized timing relations hold in in other circumstances.

Possibly of interest, the mean time for the feedback-to-tap interval for the final tap of *correct* groups was similar to that of insertion groups: 144 ms for 250 ms DAF, and 178 ms 275 ms DAF. No obvious timing measures of the correct and insertion group suggested why insertions occurred in one case but not the other.

The data for the 400p condition is even more striking. At the 400 ms delay (relatively high impairment for this tapping rate) the average feedback-to-tap time for the final tap was 268 ms, and 87% of the feedback tones started more than 200 ms before the final tap.

Although the last tap seems to be the most plausible insertion (a tap inserted

---

to cause a repetition of a tap, though this has not been explicitly tested. Tone offset or some form of "perceptual center" are other logical possibilities for the triggering event. Given the fixed delay and the 100 ms fixed length feedback tone, these will just be a fixed, constant time after feedback onset in any case.

internally to the group would need to include an explanation of how the other taps got shifted[2]), the same analysis can be applied to each of the taps. Analysis of the second tap in the 250 ms DAF case showed that 48% of the second taps in insertion groups had *no* closely preceding feedback (i.e., no sounds had occurred at all since the start of the *first* tap. In the 275 ms DAF condition 82% of the second taps had no preceding feedback. For the third tap, the mean time from the preceding feedback onset at 250 ms delay was 156 ms, with 60% of the times being greater than 200 ms; while at 275 ms delay the corresponding values were 212 ms and 84%. For the fourth tap at 250 ms delay, the values were 177 ms and 64%, and for 275 ms delay 202 ms and 64%. Thus there is no evidence for the time relation required by the Node Structure Theory.

This issue of *identifying* the insertion is a problem for tapping, one that doesn't occur in speech since units in speech have identity and the inserted unit can be precisely identified. This is a drawback to using tapping. Although a methodological problem, the extent to which unit identity itself is important to the DAF impairment is still open to question.

6.2 Insertion Group Timing: Offset Relationships

The Displaced Rhythm Hypothesis (DRH) of Howell, Powell and Khan (1983), described in Section 3.2.3, also makes a particular prediction about the timing relation between taps and feedback. In particular, impairment occurs when a signal (the auditory feedback) goes on when some other signal (here, the tap termination) goes off, and the discussion in Section 3.2.3 of speech suggests that this timing relation

[2] I.e, for tapping the easiest assumption is that 4 beats are correctly programmed and executed (with correct inter-tap timing), and then an extra beat is added at the end.

must be very close (e.g., between 0 and 50 ms).

Looking at timing relationships in the sample insertion groups (Figures 6.1 and 6.2, above) does not show any obvious support for this timing relationship, and it would be possible to provide a detailed analysis of the timing data, as was done in Section 6.1 for anticipations. However, a simpler more conceptual analysis (based on the the fixed timing between keystroke and delay) will make the necessary point. Highly impairing delays in the 250p condition of Experiment 3 are those near 250 ms. As reported in Section 5.2.3, the mean key down time (dwell) in this delay condition was about 127 ms, so the keystroke ends about 127 ms after it has been pressed. However, the feedback in this condition *always* starts 250 ms after the key is pressed. This 125 ms difference seems much too long for Howell's proposal.

The high impairment conditions for the 400p case make the case in an even stronger fashion. In this condition, DAF near 400 ms delay is maximally impairing, but the mean dwell in this condition is near 160 ms. This leaves a 240 ms gap between keystroke end and feedback on, again inconsistent with the DRH.

Another analysis confirms this point. Since key dwell in the 250p condition is about 127 ms, the DRH would predict maximal impairment near 127 ms delay, but in fact neither 100 nor 150 ms delay causes significant error impairment. Similarly, for 400 ms tapping where the dwell is 160 ms, impairment would be expected at 150 or 200 ms, but again this is not the case (see Figure 5.4).

6.3 Theoretical Discussion

Here, I will provide some concluding discussion of the two major theories discussed in this section, the Node Structure Theory and the Displaced Rhythm Hypothesis. Both these theories attempt to explain the DAF effect in a way that does not rely on a feedback guidance system. Both these theories are, however, inadequate.

6.3.1 Node Structure Theory: Discussion

In this chaper and Chapter 5, I have addressed two aspects of MacKay's explanation of DAF. One is the issue of critical interval shift, and the other relates to anticipation timing.

MacKay views the reports (see Section 2.1.2) that the delay of maximal impairment stays constant for different speech rates as important support for his theory. However, for tapping the critical interval *does* shift. The question, then, is how damaging this is to the NST. For the NST, critical interval is a direct consequence of a node's refractory period, which is hypothesized to not vary with rate changes. However, how central is this aspect of the theory? MacKay (1987) has already claimed that wefractory period can change with age; it may differ across individuals, it can change with recent activation history, and the refractory periods for different nodes (e.g., at different levels of the content hierarchy) can also be quite different. Thus, there is already a wide range of variation in refractory period timing in the theory. Since different rates of tapping would presumably be implemented via timing nodes with different rates, it is not unreasonable that refractory period might be affected by the timing period, and such an explanation might be workable.

MacKay's explanation of DAF does, however, crucially rely on the delayed auditory feedback being the stimulus for an inserted tap, and thus necessarily preceding the insertion. However, the detailed timing analysis of insertion groups in Section 6.1 shows that this does not occur. Under these circumstances, the NST cannot be a correct explanation of insertions under DAF in the tapping task.

It is, of course, possible that the results here are specific to rhythmic tapping, and do not apply to speech. However, the generality of theory would be lost, and the similarity of impairment in tapping to that of DAF in speech would be a mystery. In addition, the discussion of DAF speech data in Section 3.3.3 suggests that even the timing in speech does not allow for the feedback being a stimulus for the repetition, as the processing time for the stimulus would have to be close to 0 ms. It is certainly the case that further investigations of the detailed timing relations in actual DAF speech data would certainly be worthwhile.

The fact that the NST (as currently formulated) does not explain DAF doesn't necessary bear on the potential validity of the remainder of theory. Given that the anticipation timing results show that the NST is not an adequate explanation for DAF. the question arises as to whether there is much support for the hypothesized hyperexciteability period following the refractory period. Although the hyperexciteability period plays a small role in some other phenomena, its main purpose seems to be to account for the critical interval of DAF. Since the DAF explanation doesn't work, hyperexciteability may turn out to be dispensable in the theory. The refractory period is still certainly necessary for the NST, at least theory internally, given the bi-directional flow of activation.

6.3.2 Displaced Rhythm Hypothesis: Discussion

The Displaced Rhythm Hypothesis of Howell, Powell, and Khan (1983), as reviewed in Section 3.2.3, has two major components: (1) the particular timing requirement that feedback onset correspond with an offset, and (2) the proposal that unit identity is unimportant.

One key component of the DRH is that impairment occurs when a particular timing relation occurs, one element going on just as another goes off. Howell et al. do not provide any detailed timing analysis of speech, nor any precise quantification of how close the timing must be, but their diagrams (as well as the analysis of speech data in Section 3.2.3) suggest that it must be almost immediate. However, these relationships don't seem to apply to tapping. For the high impairment 250 ms delay condition in Experiment 3, the sound doesn't go on in most cases for an average of over 100 ms after the end (release of the tap), which is quite long when compared to the speech relationships. In addition, the conditions which would be predicted to cause most impairment (100 or 150 ms DAF, based on the average of about 125 ms dwell) is actually hardly impairing at all on an error measure.

The evidence Howell et al. provide for the DRH is based only on elapsed time measurements in error free speech, so perhaps it only applies to timing, and not to errors. But if so, it completely avoids one of the most interesting symtoms of the DAF phenomenon. Howell et al. could, of course, argue that tapping is a special case separate from speech, but they specifically intend their theory to be general (note the Morse code experiment reported in Section 2.5), and, again, this would leave the similarity of impairment across the different domains as a mystery.

Chapter 7

Experiment 4: Fixed Sequences and Interference

The preceding chapters have provided data regarding some recent theories of DAF impairment, and have argued that those theories are inadequate. Although it would be possible to develop more sophisticated feedback-based theories, it seems worth going back to one of the questions raised in the Introduction, which is whether DAF impairment should be viewed as a feedback effect at all. Perhaps all control is centrally generated, but DAF happens to create an interference pattern that (as an external, independent stimulus) disturbs the system.

Experiment 4 was designed to explore this issue. This experiment used the 4-2 pattern at a 250 ms tapping rate, and addressed the question of "feedback" versus "interference" raised in the Introduction. That is, when impairment occurs, is it because of something specific to delay, or is it simply due to the presence of a particular form of interfering pattern? Although it is hard to know what an "interfering" pattern is, the approach used was to choose fixed sequences which mimicked (to some extent) the output in an impairing delay conditions.

Experiment 4 also addressed some other issues. For reasons of space and clarity, they will not be discussed.

7.1 Experiment 4: Method

7.1.1 Subjects

18 subjects participated in this experiment.

7.1.2 Conditions/Stimuli

A total of 18 conditions were used in the 4-2 portion of Experiment 4; each condition was run 3 times, giving a total of 54 experimental trials. Only the conditions relevant to the above question will be described and reported here, although the statistics were done on the full set.

The conditions of interest almost all involved fixed sequences: sequences that were not controlled by the subjects' taps. For comparison, both SAF and 250 ms DAF conditions were also run. Regardless of whether the "feedback" sounds during the continuation phase were affected by the subjects' taps, they always had the same pitch (so that the fixed sequences "simulated" feedback).

The conditions were as follows:

1. Delay and control conditions: 0 ms delay (SAF), and 250 ms delay.

2. Fixed 4-2 sequences. These sequences continued the 4-2 pattern, although often with a phase shift. The tested conditions involved a phase shift of zero (continuing the initial metronome), 125 ms (delayed a half cycle), 250 ms (full cycle), and 500 ms (2 cycles); these will be referred to as "4-2(0)", "4-2(.5)", "4-2(1)", and "4-2(2)", respectively. An example trial for the 4-2(2) condition is shown at the top of Figure 7.1.

3. Fixed 5-3 sequence ("5-3" condition). Given that subjects typically insert beats, this condition approximately mimics the sound a subject would get in a seriously impaired trial.

4. A fixed isochronous sequence ("REG" condition). If sounds participate in triggering extra taps, then there might be many insertions here.

5. *Shifting* 4-2 sequence ("4-2(S)"). This is the most complicated one to explain, but is diagrammed at the bottom of Figure 7.1. In a true DAF condition, it is never possible to "catch up" with the sounds, because any attempt to do so will cause the sounds themselves to shift. The 4-2(S) sequence mimicked that by having a fixed 4-2 sequence, but one with a 500 ms (instead of 250 ms) pause between patterns, leading to a phase shift on *every cycle*.

6. "Random" sequences: these sequences had a non-repeating, non-predictable pattern which, like a correct 4-2 pattern, nonethelessss had 6 beats per pattern time (2 seconds). There were two random conditions: one in which the sounds corresponded with the initial underlying beat ("RAND1"), and one in which their placement was free ("RAND2"). Although the sequences were pseudo-random, the 3 trials for a condition all used the same exact sequence.

As in Experiments 2 and 3, all trials had 3 patterns of synchronization and 7 1/2 patterns of continuation, giving a total trial length of 21 seconds. The 54 trials (including some conditions not described above) gave a total tapping time of about 20 minutes.

7.1.3 Procedure

Since subjects were sometimes surprised in earlier experiments by the first difficult delay trial, an additional practice trial was included before the experimental trials

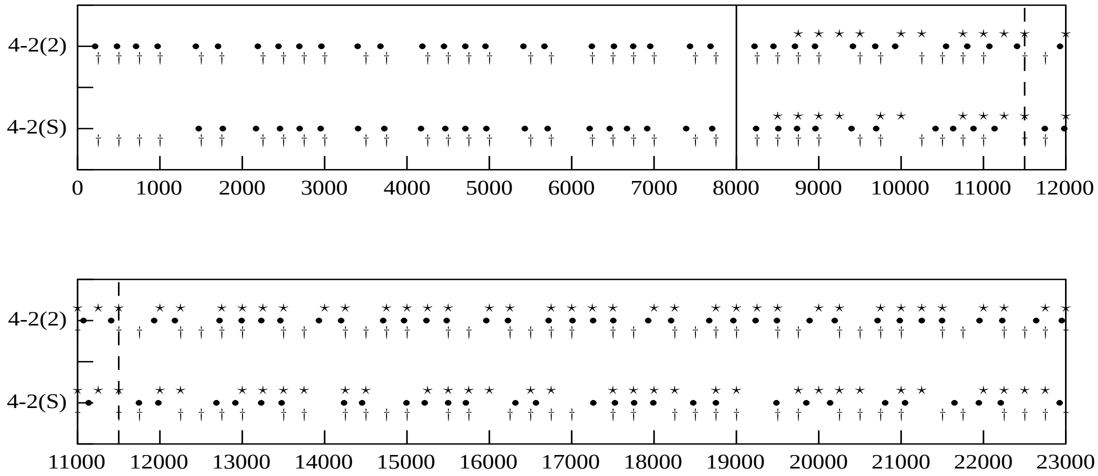


Figure 7.1: Experiment 4: Sample trials, 250p condition, 250 ms DAF. The top line is a 4-2(2) trial; in this trial, the subject followed the phase-shifted fixed pattern. The bottom is a 4-2(S) trial.

(in addition to the 3 practice sequences used in Experiments 1-3). This practice trial was a 250 ms delay trial which was not analyzed.

7.2 Experiment 4: Results

As noted earlier, the statistics here (the ANOVA and the post-hoc tests) will include all of the conditions in Experiment 4, even those not specifically described above. The insertion data for the conditions of interest are presented in Figure 7.2

Inspection of the graph shows that the 250 ms delay condition is the most impairing. The (fixed) phase-shifted 4-2 patterns are not impairing at all by this error measure. The other conditions fall between these two extremes.

Statistical analyses basically confirm these conclusions. An ANOVA over all conditions for the insertions measure was highly significant ($F(17, 289) = 11.35$, $p <$

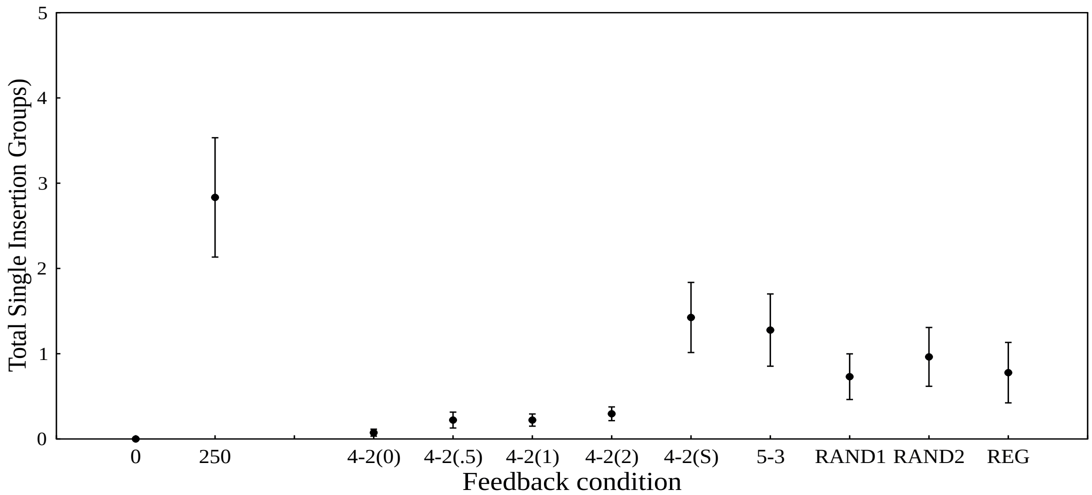


Figure 7.2: Experiment 4: Single insertion groups, fixed sequence conditions

.001). Post-hoc tests showed that 250 ms delay was the most impairing, significantly greater than all other conditions (including the 4-2(S) and 5-3 conditions). The 4-2(S) and 5-3 conditions were, however, more impairing than a number of other conditions (including the 0 ms delay condition); they did not differ significantly from the random conditions (RAND1 and RAND2). However, RAND1 and RAND2 did not differ significantly from the 0 condition. Thus, the robust differences for the primary interesting comparisons are:

```
250 > 4-2(S) = 5-3 > 0 (SAF)
```

7.3 Discussion

There are a number of different facets in this single set of results. First, it is noteworthy that 250 ms delay is robustly the most impairing condition among this set. This does not mean that further investigation might not develop a fixed sequence which was just as difficult, but for now, it seems as if there's something special about delay.

However, two other conditions also showed a significantly increased number of insertions; these were the 4-2(S) condition and the 5-3 condition. One thing which is particularly interesting is that these are the two conditions in which the groups continually shift forward in phase relative to the original grouping. This occurs in the 4-2(S) condition because it was explicitly designed that way, and the shift is also implicit in the 5-3 condition because of the fact the group is longer (i.e., it is a 10 beat pattern, not an 8 beat pattern).

The random conditions did not show impairment significantly different from the SAF condition, although they do show a trend towards impairment (and they might show statistically reliable differences in a more focussed experiment).

There are two interesting interpretions from this data. First, the (relative) absence of impairment with random sequences suggests that some sort of pattern may be necessary to see insertion impairment. More importantly, the two fixed patterns which cause significant impairment (4-2(S) and 5-3) are the ones that shift forward in time. This suggests the hypothesis that subjects are trying to keep up with the stimuli, and that they insert taps in the process of doing so. This idea will be developed in more detail in Chapter 8.

Chapter 8

Entrainment

None of the theories reviewed so far (feedback guidance, the Displaced Rhythm Hypothesis, or the Node Structure Theory) seems to be adequate; the DAF phenomenon is still unexplained. I would like to take the result of Chapter 7 (where insertions occurred in those fixed patterns that shifted in time relative to the subject's taps), combine that with some other results, and use that as the starting point for a new class of hypotheses. These hypotheses claim that the impairment under DAF (insertions in particular) is due to subjects trying to synchronize with the delayed stimulus. In other words, subjects are trying to match their taps with the anticipated sounds.

I will refer to this approach as an "entrainment" hypothesis, using entrainment in a somewhat informal sense. Entrainment has a number of more specific meanings, usually applying to the mathematical or physical coupling of two oscillators.[1]

---

[1]It is non-trivial to apply existing entrainment proposals in the dynamical systems literature to the current situation, using mathematically defined constructions such as (continuously) coupled Van der Pohl oscillators (e.g., that the stimulus (here delayed) acts as a forcing signal to the oscillator generating the taps). The primary problem is that the pattern is "gapped". I do not know of specific for *generating* gapped patterns in the dynamical systems literature, although there are proposals

Here, it refers to an external pattern which is inducing a subject's internal pattern generator to follow or synchronize with it.

Why consider such an approach? The fixed sequence data of Chapter 7 are one motivation. Also, when subjects in these tapping experiments were asked how they coped with the delayed condition, they often reported a strategy of "trying to keep up with the sound". There is extensive anecdotal evidence that *not* entraining to a rhythmic stimulus can be hard: Lashley (1951) uses the example of rhythmic music tending to induce people to walk in synchrony with it.

As will be seen, some further supportive evidence will come from two sources:

1. Evidence from Experiment 3 that the timing of insertion taps is quite consistent and regular, suggesting the presence of an underlying oscillator.

2. Evidence from Experiment 4 that subjects who tend to entrain to rhythmic patterns are also those who are most highly impaired in DAF conditions.

Although such an entrainment hypothesis has not been explicitly proposed before, it has sometimes been foreshadowed (although not in any detail) by earlier work. Fairbanks and Guttman (1958) refer to the possibility of DAF impairment being due to the attempt to restore "normal phase relationships" between feedback and output. On a more general issue, Chase et al (1961b) argue that "the sensory feedback systems used for monitoring temporally complex motor events are easily disturbed by sensory events of any type which are displaced in time from the motor events being monitored." Yates (1963b) suggested that "deviation of a marked kind from the *expected* pattern of feedback will lead to an attempt to restore the usual

---

for *perception* of such patterns, and for entraining oscillators *to* such patterns. For example, Large and Jones (1998) provide an (attentional) tracking model, in which multiple coupled oscillators (at different frequencies) track different components of a gapped (and slowly shifting) pattern. Interestingly, the coupling of the *perceiving* oscillators in this model is actually quite discrete.

relationship" (italics added), where the "expected pattern" of feedback is built up as a result of a vast amount of experience. This entrainment proposal is also in line with the *general* sense of Howell et al.'s Displaced Rhythm Hypothesis, although not with the specifics. For example, Howell et al. (1983) provide a a very specific hypothesis about offset timing relationships, but also state in more general terms that "A rhythmic signal with a similar pattern to the activity being organized but out of synchrony with that action is particularly effective in disturbing performance".

The following section will argue that the insertion data of Experiment 3 supports the proposal that there is an underlying oscillator. The next section will describe in more detail how entrainment might account for the pattern of DAF impairment, and will provide the outlines of a model. The final section will then provide some evidence that subjects *are* entraining to the delayed stimuli.

8.1 Experiment 3 Analysis: Underlying Oscillator

One striking fact is that inserted taps are quite close in time with the underlying beat. This point is best illustrated by example. Figure 8.1 contains some sample trials for Experiment 3 (in the same format as Figure 4.1, which makes the same point). The insertions appear to be in time; they are not simply inserted at random intervals. Some feedback theories would suggest that the insertions will occur somewhat randomly, depending on when feedback is received (e.g., in response to an unexpected "error" signal). An oscillator-based pattern generation mechanism would, however, predict that all taps (including insertion taps) will fall on the underlying beat.

An analysis of this can be done in detail, looking at the WGI for each serial

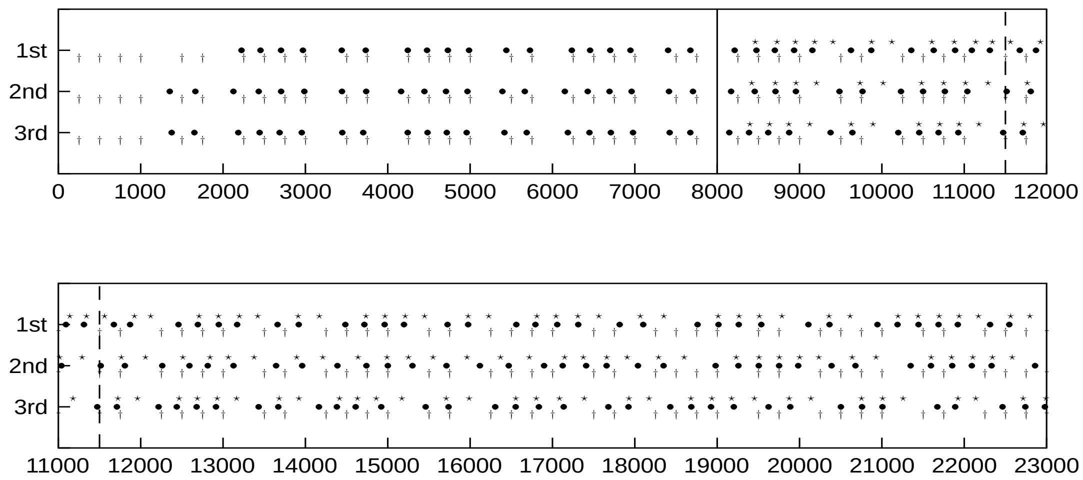


Figure 8.1: Experiment 3: Sample trials, 250p condition, 250 ms DAF

.

position in all insertion groups (for simplicity, only looking at the 4-tap group). Assume, tentatively, that the insertion tap is the final tap of a 5-tap insertion group. Looking at insertion groups and averaging over all conditions, the mean WGIs for each position are as follows: first position: 262 ms, second position: 249 ms, third position: 238 ms, and final (fourth) position: 247 ms. Restricting the analysis to the highly impairing 250 ms DAF condition, the corresponding insertion group times are 260 ms, 248 ms, 236 ms, and 251 ms.

At one level, the relevant data here is that none of the mean times in insertion groups departs far from the nominal 250 ms tapping rate; no more than 14 ms. This seems more consistent with an underlying oscillator model than with an error-correcting model.

However, it should also be noted that an ANOVA on the mean insertion group WGI times, with serial position as the factor, *is* significant ($F(3, 51) = 12.89$, p <

.001). Of the four times (262, 249, 238, and 247 ms), a Newman-Keuls test showed that the first WGI (262 ms) was higher than all the others, and that the third WGI (238 ms) was less than all the others. Thus, it is not the case that there are no differences between the different serial positions.

There are two points of interest regarding this data. First, the longest WGI appears to be the initial one (that is, between the first tap and the second tap); this might be taken as an indication of an attempt to wait and delay the second tap until the sound arrives from the first tap. The other point is that careful interpretation of these data will require a more thorough timing analysis of serial WGI effects in group tapping in general, since there may well be phenomena such as an implicit accent on the first beat of the group which would contribute to small timing differences, whether an insertion had taken place or not.

Thus, it seems fair to conclude that the timing within groups is relatively even, adhering to an underlying clock.

8.2 Impairment Explanation and Model

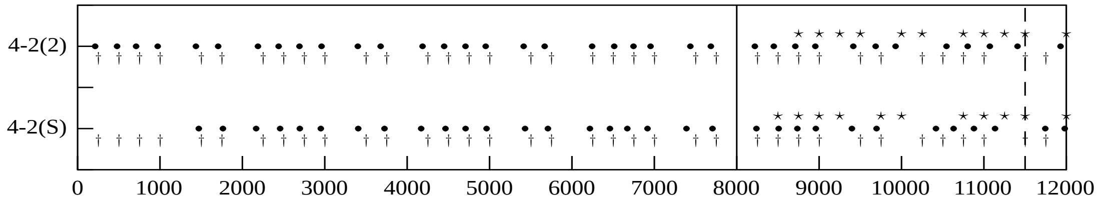
How might entrainment explain the symptoms of DAF impairment? Assume that pattern generation of the sort of rhythmic patterns used here is based on generating the correct *gaps* on top of a stable underlying oscillator, and assume that the subject is trying to keep up with the delayed signal. The beat oscillator is fixed, but the oscillators which generate the gaps (tap *absences*) can shift when there is evidence of a mismatch between sound and tap. Since there's a strong underlying oscillator, it is only possible to shift in a way that maintains that basic timing structure; that is, shifting will only occur when the entraining pulses match the internal generator pulses. Since the DAF sounds are *ahead* of the taps, it is necessary to shift forward to stay on the delayed beat. This forward shift could occur via insertion of an extra tap, or by shifting forward the gap-generating oscillators. A simple model of this latter approach can be constructed which generates the 5-3 pattern characteristic of severe DAF insertion on the 4-2 pattern. Of course, any such shift also causes the stimulus train to shift forward one unit, resulting in a vicious cycle in which it is impossible to catch up.

The underlying (non-perturbable) beat oscillator is thus responsible for the critical interval, and the attempt to match the gapped pattern on top of this oscillator is responsible for the insertions.

8.3 Experiment 4: Entrainment Correlations

The most striking evidence in support of an entrainment proposal comes from a post-hoc analysis on some data of Experiment 4. Observation of different subjects' performances in Experiment 4 suggested one possible source for subjects' differences in impairment in the DAF condition. Experiment 4 included, in addition to DAF conditions, conditions in which subjects heard fixed sequences that did not change in response to subjects' taps. Among these fixed sequences were ones which continued the pattern and rate set up by the pacing signal, but sometimes with a shifted phase (i.e., offset in time from the pacing signal). The pitch of these sequences was, however, the same pitch as the feedback tone. What was observed was that some subjects would tend to switch their tapping to be in synchrony with the fixed phase-shifted signal (as illustrated in the top line of 8.2, copied from 7.1). The question this raises is whether subjects who engage in following the pattern in this fashion are the same subjects who are highly impaired under DAF. If so, this would be consistent with the proposal that DAF impairment is due to subjects trying to "keep up" with the delayed signal.

To test this possibility, a measure was done on the extent to which each subject



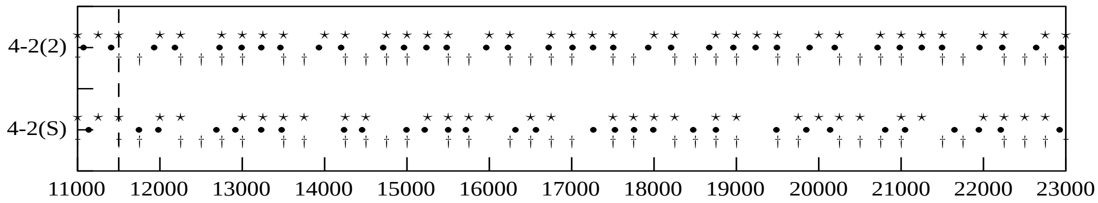


Figure 8.2: Experiment 4: Sample trials, 250p condition, 250 ms DAF. The top line is a 4-2(2) trial; in this trial, the subject followed the phase-shifted fixed pattern. The bottom is a 4-2(S) trial.

synchronized with the fixed pattern. This measure was then run in a correlation with the (total) insertion error count in the high impairment conditions (i.e., with 250 ms and 275 ms delay).

For the measure of "synchronization", 2 judges (including the author) were provided with graphs in which subjects' taps and the tones were graphed on the same time line (as in Figure 8.2). Such graphs were prepared for three of the fixed 4-2 patterns of Experiment 4 (4-2(0), 4-2(1) and 4-2(2)) The judges' task was to count (for each such trial) how many *full* 4-2 patterns of the continuation phase had subjects' taps and "feedback" tones in synchrony. The judges were given some extreme examples of obvious synchrony and asynchrony, but were otherwise expected to set their own criterion level. The ratings of the two judges had a correlation of .99, suggesting that such ratings are fairly obvious and consistent. The mean of the two judges' ratings for a trial was used as the sychronization measure.

Correlations were then done combining this synchronization measure (per subject), and that subject's impairment (single insertion error count) for the impairing DAF conditions of Experiment 4. The correlations involving the 4-2(2) synchronization measure were robustly quite high. The correlation with impairment in the 250 ms DAF condition was $r = .76$, $p < .001$; for the 275 ms DAF condition, it was $r = .73$, $p < .001$. Correlation of this synchronization with impairment in three other (unreported) conditions in a second part of Experiment 4 were also quite high, all above .7, with p values less than .005. The correlations using the 4-2(1) synchronization measure were less reliable, but still suggestive: for 250 ms delay, $r = .47$, $p < .05$; for 275 ms delay, $r = .43$, $p = .07$.

In other words, subjects who tended to phase-shift to match a shifted 4-2 pattern were also those who had high error rates in the DAF condition. Such correlations might help explain the variety of individual reponses to DAF. For speech, there have been repeated reports of wide variation in the DAF effect, with some subjects showing extreme impairment, and some subjects being almost unaffected by it. This was true in the current tapping experiments as well. The correlation suggests that DAF susceptibility may be related to a more general entrainment susceptibility.

8.4 Discussion

Such a theory could be extended to speech, although this possibility has not been worked out in detail. Speech is clearly rhythmical, based on syllables or stress (see, e.g., Martin, 1972). In the case of DAF, *rhythmic* entrainment might occur on the basis of the rhythm of the delayed sound. This proposal bears some relation to the predictive, syllable-based speech proposal of Harrington (1987, 1988).

There is one possible issue with accepting this as a general explanation for DAF impairment. The experiments here used a continuation paradigm, in which each trial started out with subjects *explicitly* entraining to a pacing tone; perhaps the correlation above is just an artifact of this experimental method. However, as has been repeated emphasized, the results in the current experiments fundamentally replicate existing results which have not used a continuation paradigm, so this explanation seems unlikely. It is true, however, that this possibility should be checked by rerunning these experiments with a procedure which does not use an explicit auditory synchronization phase.

Clearly, the argumentation here in support of an entrainment hypothesis is preliminary. However, the correlation data are quite striking, and suggest that these sorts of proposals should be looked at more closely.

Chapter 9

Conclusions

In this concluding section, I will first summarize the empirical results of this thesis, and their theoretical import. I will then discuss two of the questions raised in the Introduction in light of this new data: the issue of whether DAF in speech and tapping is the same phenomenon, and the issue of whether the DAF effect relates to feedback guidance, or is an interference effect.

This dissertation is the first study of DAF and tapping to make use of current computer technology; this has allowed much more extensive and precise data collection and analysis than previous studies. In agreement with earlier studies (e.g., Chase et al., 1961b; Ruhm and Cooper, 1962), robust impairment been found with DAF and patterned tapping. In accordance with earlier anecdotal reports of insertions in tapping, such insertions have been quantitatively established as the dominant form of counting error. The temporal control allowed by the continuation paradigm has also allowed demonstration of a critical interval (delay of maximal impairment) in tapping, as has been reported for speech. However, in contrast to some claims for speech, this critical interval is not fixed, but depends on the tapping rate, with the critical interval in fact being almost identical to the tapping rate itself. This coincidence of tap and (delayed) sound seems likely to be central to an explanation of DAF impairment for tapping, and, by extension, possibly for speech.

Further analyses of the relative proportion of insertions and total errors has established that the proportion of insertion errors in tapping is maximal near the critical interval, corresponding to the pattern reported by Fairbanks and Guttman (1958) for speech.

Timing analyses of the relationships between feedback sounds and taps were investigated from the perspective of two particular theories, the Node Structure Theory of MacKay (1987) and the Displaced Rhythm Hypothesis of Howell et al. (1983). The data were not consistent with the predictions of either theory.

Thus, no existing theories account for the DAF impairment data. Based on data showing that certain fixed but phase-shifting patterns caused noticeable impairment, a theory involving entrainment was advanced, and evidence from high correlations between individual's tendencies towards entrainment and their impairment under DAF was provided in support of this hypothesis.

What do the data say about the relationship between DAF effects in speech and tapping? Two additional aspects of DAF tapping impairment have been found which correspond to impairment in speech: the presence of a critical interval, and the maximal proportion of insertion errors at that interval. These findings certainly strengthen the view that the same phenomenon is at work in the two cases. If this similarity holds, there are two important conclusions to be drawn about the possible causes of DAF impairment:

1. Explanations for DAF impairment in speech have sometimes relied on the conflict between delayed air-conducted auditory feedback and synchronous bone-conducted feedback. However, the tapping experiments here explicitly removed any component of synchronous auditory feedback (something not feasible in speech). If the same mechanism is at work in the two cases, this argues that this particular feedback conflict is not necessary for DAF impairment.

Proprioceptive feedback was *not* altered in the current tapping experiments, so feedback conflict may certainly be part of an explanation. However, such explanations will need to deal with how information from fundamentally different feedback information sources is combined.

2. Explanations for DAF impairment in speech sometimes rely on specific aspects of speech articulation, or on the identity of units (phonemes, syllables) in speech. However, individual taps do not have such identity, but similar impairment nonetheless occurs. Again, if one takes the strong view that the same mechanism is active in the two cases, this suggests that unit identity is not central to the DAF effect.

Of course, some differences from reported speech results have also been found. For instance, no reliable key force increase (corresponding roughly to speech loudness) was found in the current experiments; some possible reasons for this were discussed in Section 4.4.3. Perhaps more significantly, the tapping data show a shift in critical interval with tapping rate, whereas a number of studies have argued that this does not occur for speech. However, in light of the small number of these studies, and the presence of some conflicting data (see Section 2.1.3), the current finding suggests that this should be looked at more carefully for speech.

There are some disadvantages to using single-finger tapping as a domain of study. The task is simple and controllable, but without unit identification it is impossibly to definitively identify exactly which tap is the "incorrect" insertion, something which is fairly obvious for speech. Similarly, in the absence of distinguishable units it is not possible to look at issues of perseveration and anticipation (in the sense of Dell et al., 1997, or Palmer and van de Sande, 1995), since each tap is effectively identical. One approach to addressing this weakness would be to extend the tapping studies to use multiple fingers, with each finger producing a different tone. Of course, such motor/sound correspondences will be necessarily much weaker than those for speech; skilled musical instrument performance might be a good intermediate domain.

The remaining general question is whether the impairing effect of DAF is due to its being processed as feedback by a feedback guidance system, or is due to some form of interference. Although two specific theories which have argued for a non-feedback based explanation (the Node Structure Theory and the Displaced Rhythm Hypothesis) have been shown to be inadequate, there is also no feedback theory which explains the data. Thus, the question is still open. However, Experiment 4 provided some data which showed that some particular fixed sequences (those that involved a continuous phase shift over time) were significantly impairing, leading to a preliminary proposal based on entrainment. Entrainment, of course, is not a feedback based explanation, but rather one in which an external pattern *similar to* the generated pattern is capable of causing the generated pattern to become similar to the entraining pattern.

It is perhaps impossible to prove either a feedback or an interference proposal. Internally to the system, the differences seems straightforward: in a feedback system, the stimuli are processed in relation to an internal model of correct output, and the stimuli are experienced or interpreted as output of the system itself. In an interference proposal, the stimuli are independent of the system and are generated externally, but happen to be able to perturb the system in a particular manner. However, the boundaries get blurred: if the entrainment proposal is correct, then in the DAF case a stimulus which is in fact generated by the subject (albeit temporally shifted) is being processed as an external stimulus. The opposite might also occur: an external, independent stimulus might be interpreted as feedback.

This distinction may not be definitively resolvable, since for any instance of feedback-based impairment, it could be claimed that the same impairment would have resulted if the subject had heard that identical stimulus sequence (perhaps also requiring identical overlap with the motor acts), even if it had not actually been produced by the subject. A stimulus per se (as a physical manifestation in the world) is not given an identification tag which states whether it is internally or externally generated (although it may be identifiable in some cases, e.g., one's own speech). Establishing how such associations are made is likely to be a very deep question.

Chapter 10

References

Banton, L.J. 1995. The role of visual and auditory feedback during the sight-reading of music. *Psychology of Music*, **23**, 3-16.

Black, J.W. 1951. The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, **16**, 56-60.

Bootsma, R.J, and van Wieringen, P.C.W. 199. Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 21-29.

Borden, G.J. 1979. An interpretation of research on feedback interruption in speech. *Brain and Language*, **7**, 307-319.

Bunker, J.W.M. 1912. *The distribution of the colon bacillus in polluted oysters.* PhD Dissertation, Brown University.

Burke, B.D. 1975. Susceptibility to delayed auditory feedback and dependence on auditory or sensory feedback. *Journal of Communication Disorders*, **8**, 75-96.

Butler, R.A., and Galloway, F.T. 1957. Factoral analysis of the delayed speech feedback phenomenon. *Journal of the Acoustical Society of America*, **29**, 632-635.

Carlton, L.G. 1981. Processing visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1019-1030.

Chase, R.A. 1958. Effect of delayed auditory feedback on the repetition of speech sounds. *Journal of Speech and Hearing Disorders*, **23**, 583-590.

Chase, R.A. 1965. An information-flow model of the organization of motor activity. I: Transduction, transmission, and central control of sensory information. *Journal of Nervous and Mental Disease*, **140**, 239-251.

Chase, R.A., and Guilfoyle, G. 1962. Effect of simultaneous delayed and undelayed auditory feedback on speech. *Journal of Speech and Hearing Research*, **5**, 144-151.

Chase, R., Harvey, S., Standfast, S., Rapin, I., and Sutton, S. 1959. Comparison of the effects of delayed auditory feedback on speech and key tapping. *Science*, **129**, 903-904.

Chase, R., Harvey, S., Standfast, S., Rapin, I., and Sutton, S. 1961(a). Studies on sensory feedback I: Effect of delayed auditory feedback on speech and keytapping. *Quarterly Journal of Experimental Psychology*, **13**, 141-152.

Chase, R., Rapin, I., Gilden, L, Sutton, S., and Guilfoyle, G. 1961(b). Studies on sensory feedback II: Sensory feedback influences on keytapping motor tasks. *Quarterly Journal of Experimental Psychology*, **13**, 153-167.

Chase, R.A., Sutton, S., Fowler, E.P., Fay, T.H., and Ruhm, H.B. 1961(c). Low sensation level delayed clicks and keytapping. *Journal of Speech and Hearing Research*, **4**, 73-78.

Chase, R.A., Sutton, S., and Rapin, I. 1961(d). Sensory feedback influences on motor performance. *Journal of Auditory Research*, **3**, 212-223.

Cowie, R.I.D, and Douglas-Cowie, E. 1983. Speech production in profound postlingual deafness. In *Hearing Science and Hearing Disorders*, M.E. Lutman and M.P.Haggard (Eds). London: Academic Press.

Delcomyn, F. 1980. Neural basis of rhythmic behavior in animals. *Science*, **210**, 492-498.

Dell, G.S, Burger, L.K., and Svec, W.R. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review*, **104**, 123-147.

Fabbro, F., and Daro, V. 1995. Delayed auditory feedback in polyglot simultaneous interpreters. *Brain and Language*, **48**, 309-319.

Fairbanks, G. 1954. Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, **19**, 133-139.

Fairbanks, G. 1955. Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders*, **20**, 333-346.

Fairbanks, G. 1958. Effects of delayed auditory feedback upon articulation. *Journal of Speech and Hearing Research*, **1**, 12-22.

Fillenbaum, S. 1963. Impairment in performance with delayed auditory feedback as related to task characteristics. *Journal of Verbal Learning and Verbal Behavior*, **2**, 136-141.

Fillenbaum, S. 1965. Adaptation with constant and variable delay in auditory feedback. *Psychonomic Science*, **3**, 45-46.

Fillenbaum, S, and Wiesen, R. 1961. Contextual constraints and disruptions in reading with delayed auditory feedback. *Journal of the Acoustical Society of America*, **33**, 1800-1801.

Finney, S.A. 1997. Auditory feedback and musical keyboard performance. *Music Perception*, **15**, 153-174.

Fletcher, J.F., and Yates, A.J. 1971. The repetition of single speech sounds under delayed auditory feedback. *British Journal of Psychology*, **62**, 73-79.

Fraisse, P, and Voillaume, C. 1971. Les reperes du sujet dan la synchronisation et dans la pseudo-synchronisation. *L'annee Psychologique*, **71**, 359-369.

Garber, S.R, and Siegel, G.M. 1982. Feedback and motor control in stuttering. In *Learning, speech, and the complex effects of punishment*, Routh, D.K. (ed), Plenum Press, NY.

Garner, W.R. 1974. *The Processing of Information and Structure.* Lawrence Erlbaum, Potomac, MD.

Gates, A., and Bradshaw, J. 1974. Effects of auditory feedback on a musical performance task. *Perception and Psychophysics*, **16**, 105-109.

Gates, A., Bradshaw, J., and Nettleton, N. 1974. Effect of different delayed auditory feedback intervals on a music performance task. *Perception and Psychophysics*, **15**, 21-25.

Goldiamond, I., Atkinson, C.J., and Bilger, R.C. 1962. Stabilization of behavior and prolonged exposure to delayed auditory feedback. *Science*, **135**, 437-438.

Greeno, J.G., and Simon, H.A. 1974. Processes for sequence production. *Psychological Review*, 81, 187-198.

Hanley, C.N., and Tiffany, W.R. 1954. An investigation into the use of electomechanically delayed side tone in auditory testing. *Journal of Speech and Hearing Disorders*, **19**, 367-374.

Harrington, J. 1987. A model of stuttering and the production of speech under delayed auditory feedback conditions. In *Speech Motor Dynamics in Stuttering*, Peters, H.F.M, and Hulstijn, W. (eds), Springer-Verlag, Vienna.

Harrington, J. 1988. Stuttering, delayed auditory feedback, and linguistic rhythm. *Journal of Speech and Hearing Research*, **31**, 36-47.

Havlicek, L. L. 1968. Effects of delayed auditory feedback on musical performance. *Journal of Research in Music Education*, **16**, 308-318.

Howell, P. 1983. The effect of delaying auditory feedback of selected components of the speech signal. *Perception and Psychophysics*, **34**, 387-396.

Howell, P., and Archer, A. 1984. Susceptibility to the effects of delayed auditory feedback. *Perception and Psychophysics*, **36**, 296-302.

Howell, P., El-Yaniv, N., and Powell, D.J. 1987. Factors affecting fluency in stutterers when speaking under altered auditory feedback. In *Speech Motor Dynamics in Stuttering*, Peters, H.F.M, and Hulstijn, W. (eds), Springer-Verlag, Vienna.

Howell, P., and Powell, D.J. 1984. Hearing your voice through bone and air: Implications for explanations of stuttering behavior from studies of normal speakers. *Journal of Fluency Disorders*, **9**, 247-264.

Howell, P., and Powell, D.J. 1987. Delayed auditory feedback with delayed sounds varying in duration. *Perception and Psychophysics*, **42**, 166-172.

Howell, P., Powell, D.J., and Khan, I. 1983. Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 772-784.

Huggins, A.W.F. 1968. Delayed auditory feedback and the temporal properties of the speech material. *Zeitschrift fur Phonetik, Sprachwissenscahft, und Kommunikations-Forschung*, **21**, 53-60.

James, W. 1890. *Psychology.* Henry Holt and Company, NY.

Jones, M.R. 1981. A tutorial on some issues and methods in serial pattern research. *Perception and Psychophysics*, **30.5**, 492-504.

Kalmus, H, Denes, P, and Fry, D.B. 1955. Effect of delayed acoustic feedback on some non-vocal activities. *Nature*, **175**, 1078.

Karlovich, R.S, and Graham, J.T. 1966. Effects of pure tone synchronous and delayed auditory feedback on keytapping performance to a programmed visual stimulus. *Journal of Speech and Hearing Research*, **9**, 596-603.

Karlovich, R.S, and Graham, J.T. 1967. Visually and auditorily paced keytapping performance during synchronous, decreased, and delayed auditory and visual feedback. *Journal of Speech and Hearing Research*, **10**, 859-864.

Karlovich, R.S, and Graham, J.T. 1968. Auditorily paced keytapping performance during synchronous, decreased, and delayed visual feedback. *Perceptual and Motor Skills*, **26**, 731-743.

Keele, S., and Posner, M. 1968. Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, **77**, 155-158.

Keele, S. 1987. Sequencing and timing in skilled perception and action. In *Language Perception and Production*, Allport, A., MacKay, D., Prinz, W., and Scheerer, E. (eds.), Academic Press, London.

Kelso, J.A.S. 1995. *Dynamic Patterns.* MIT Press, Cambridge.

Klein, G.S., and Wolitzky, D.L. 1970. Effects of occluding auditory feedback from one's own voice. *Journal of Abnormal Psychology*, **75**, 50-56.

Kodman, F. 1961. Controlled reading rate under delayed speech feedback. *Journal of Auditory Research*, **1**, 186-193.

Kozhenikov, V.A., and Chistovich, L.A. 1965. *Speech: Articulation and Perception.* Clearinghouse for Federal Scientific and Technical Information: Washington, DC.

Lane, H., and Tranel, B. 1971. The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, **14**, 677-709.

Large, E.W., and Jones, M.R. 1998. The dynamics of attending: How we track time-varying events. In Press, *Psychological Review* .

Lashley, K. 1951. The problem of serial order in behavior. In *Cerebral Mechanisms of Behavior*, L.A. Jeffress (Ed). New York: Wiley.

Laszlo, J.I. 1967. Kinaesthetic and exteroceptive information in the performance of motor skills. *Physiology and Behavior*, **2**, 359-365.

Leder, S.B., Spitzer, J.B., Kirchner, J.C, Flevaris-Phillips,C., Milner, P., and Richardson, F. 1987. Speaking rate of adventitiously deaf male cochlear implant candidates. *Journal of the Acoustical Society of America*, **82**, 843-846.

Lee, B.S. 1950(a). Some effects of side-tone delay. *Journal of the Acoustical Society of America*, **22**, 639-640.

Lee, B.S. 1950(b). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, **22**, 824-826.

Lee, B.S. 1951. Artificial stutter. *Journal of Speech and Hearing Disorders*, **16**, 53-55.

Lenneberg, E.H. 1967. *Biological Foundations of Language.* Wiley, New York.

MacKay, D.G. 1968. Metamorphosis of a critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *Journal of the Acoustical Society of America*, **43**, 811-821.

MacKay, D.G. 1969. To speak with an accent: Effects of nasal distortion on stuttering under delayed auditory feedback. *Perception and Psychophysics*, **5**, 183-188.

MacKay, D.G. 1970. How does language familiarity influence stuttering under delayed auditory feedback? *Perceptual and Motor Skills*, **30**, 655-669.

MacKay, D.G. 1982. The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review*, **89**, 483-506.

MacKay, D.G. 1985. A theory of the representation, organization, and timing of action with implications for sequencing disorders. In *Neuropsychological Studies of Apraxia*, Roy, E.A. (ed), Elsevier.

MacKay, D.G. 1987. *The organization of perception and action.* Springer-Verlag, New York.

MacKenzie, C.L, and Marteniuk, R.G. 1985. Motor skill: Feedback, knowledge, and structural issues. *Canadian Journal of Psychology*, **39.2**, 313-337.

Martin, J.G. 1972. Rhythmic (hierarchical) vs serial structure in speech and other behavior. *Psychological Review*, **79**, 487-509.

Mates, J., Radil, T., and Poppel, E. 1992. Cooperative tapping: Time control under different feedback conditions. *Perception and Psychophysics*, **52**, 691-704.

Massey, H. 1987. *The Complete DX7II.* New York: Amsco.

Palmer, C., and van de Sande, C. 1995. Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 947-962.

Rapin, I., Costa, L.D., Mandel, I.J., and Fromowitz, A.J. 1963. Effect of varying delays in auditory feedback on key-tapping of children. *Perceptual and Motor Skills*, **16**, 489-500.

Rapin, I., Costa, L.D., Mandel, I.J., and Fromowitz, A.J. 1966. Keytapping and delayed feedback. *Journal of Speech and Hearing Research*, **9**, 278-288.

Repp, B. 1998. Effects of auditory feedback deprivation on expressive piano performance. Manuscript submitted for publication.

Restle, F. 1970. Theory of serial pattern learning: Structural trees. *Psychological Review*, **77**, 481-495.

Robinson, D.O. 1970. The limits of selective attention in DAF shadowing. *Psychonomic Science*, **21**, 325-327.

Robinson, G.M. 1972. The delayed auditory feedback effect is a function of speech rate. *Journal of Experimental Psychology*, **95**, 1-5.

Roehrig, W.C. 1965. Addition of controlled distortion to delay of auditory feedback. *Perceptual and Motor Skills*, **21**, 407-413.

Rosenbaum, D., Kenny, S., and Derr, M. 1983. Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 86-102.

Ruhm, H.B., and Cooper, W.A. 1962. Low sensation level effects of pure-tone delayed auditory feedback. *Journal of Speech and Hearing Research*, **5**, 185-193.

Ruhm, H.B., and Cooper, W.A. 1963. Some factors that influence pure tone delayed auditory feedback. *Journal of Speech and Hearing Research*, **6**, 207-222.

Ruhm, H.B., and Cooper, W.A. 1964. Influence on motor performance of simultaneous delayed and synchronous pure-tone auditory feedback. *Journal of Speech and Hearing Research*, **7**, 175-182.

Semjen, A., Vorberg, D. and Schulze, H-H. 1998. Getting synchronized with the metronome: Comparisons between phase and period correction. *Psychological Re-*

search, **61**, 44-55.

Siegel, G.M., Fehst, C.A., Garber, S.R., and Pick, H.L. 1980. Delayed auditory feedback with children. *Journal of Speech and Hearing Research*, **23**, 802-813.

Siegel, G.M., Schork, E.J., Pick, H.L., and Garber, S.R. 1982. Parameters of auditory feedback. *Journal of Speech and Hearing Research*, **25**, 473-475.

Smith, K.U. 1962. *Delayed sensory feedback and behavior.* W.B. Saunders, Philadedelphia.

Spilka, B. 1954(a). Some vocal effects of different reading passages and time delays in speech feedback. *Journal of Speech and Hearing Disorders*, **19**, 37-47.

Sullivan, R.F., Polisar, I.A., Ruffy, R.L, and Miller, J.H. 1973. Relative influence of the Lombard and delayed auditory feedback tests for functional hearing loss on phonation/time ration. *Journal of Auditory Research*, **13**, 50-63.

Sussman, H.M., and Smith, K.U. 1971. Jaw movements under delayed auditory feedback. *Journal of the Acoustical Society of America*, **50**, 685-691.

Toates, F.M. 1975. *Control Theory in Biology and Experimental Psychology.* Hutchinson Educational, London.

Vaughn, J., Mattson, T.R., and Rosenbaum, D.A. 1998. The regulation of contact in rhythmic tapping. In *Timing of Behavior*, Rosenbaum, D.A., and Collyer, C.E. (eds), MIT Press: Cambridge.

Voillaume, C. 1971. Models pour l'etude de la regulation des mouvents cadences. *L'annee Psychologique*, **71**, 347-358.

Waldstein, R. 1990. Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America.*, **88**, 2099-2114.

Wiener, N. 1948. *Cybernetics*. MIT Press, Cambridge, MA.

Wing, A.M. 1977. Perturbations of auditory feedback delay and the timing of movement. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 175-186.

Wing, A.M., and Kristofferson, A.B. 1973. The timing of interresponse intervals. *Perception and Psychophysics*, **13**, 455-460.

Yates, A.J. 1963 (a). Delayed auditory feedback. *Psychological Bulletin*, **60**, 213-232.

Yates, A.J. 1963 (b). Recent empirical and theoretical approaches to the experimental manipulation of speech in normal subjects and in stammerers. *Behaviour Research (and?) Therapy*, **1**, 95-119.

Yates, A.J. 1965 (a). Effects of delayed auditory feedback on Morse code transmission by skilled operators. *Journal of Experimental Psychology*, **69**, 467-473.

Yates, A.J. 1965(b). Delayed auditory feedback and shadowing. *Quarterly Journal of Experimental Psychology*, **17**, 125-131.

Zimmerman, G., Brown, C., Kelso, J.A.S., Hurtig, R., and Forrest, K. 1988. The association between acoustic and articulatory events in a delayed auditory feedback paradigm. *Journal of Phonetics*, **16**, 437-451.

Appendix A

Experimental Instructions

The following are the written instructions used in the tapping experiments:

In this experiment, you will tap on a keyboard in response to sounds you hear over headphones; you will also sometimes hear sounds in response to your taps. Each trial will last about 25 seconds, and will consist of the following sequence of events:

1. First, a loud hiss ("white noise") will come over the headphones.

2. Almost immediately, a sequence of pacing sounds will start, which will be a simple pattern; within each block of trials, the pattern will always be the same. As soon as possible, start tapping along with the sounds. You will hear no sounds in response to your taps. This is the "synchronization" phase.

3. After about 10 seconds, the pacing sounds will stop, but new sounds (at a different pitch) may start; these may occur at the same time as your taps,

or may not. Whatever the sounds, your task is to try to continue to tap at the same speed and pattern as you did during the pacing signals. This is the "continuation" phase.

4. The trial is over when the white noise stops.

Sometimes the sounds during the continuation phase may make it difficult to tap regularly. You should try to ignore the sounds if this occurs; try to keep tapping regularly no matter what you hear.

You will not be able to see the keyboard during the experiment. Always tap with the index finger of your dominant hand (i.e., if you are right-handed, with your right hand), tap from your wrist, and try not to lift your finger off the key during a trial. During the synchronization phase, it is best if you try to make the key hit the bottom at the same time as the pacing sound.

After each trial, the experimenter will wait for you to say "Go ahead" or "Ready" before proceeding to the next trial.

You will be tapping for a while, so please be conscious of any fatigue in your hand or wrist, and take a rest break at any point for any reason. The actual tapping in the experiment totals about 30 minutes; the experiment as a whole will take 50-60 minutes with breaks.

Appendix B

Data Analysis Details

This section will describe some details and issues involving experimental design, implementation, and analysis which seemed out of place in the main text. For various reasons, this Appendix is a bit rougher and more informal than the body of the text, but I feel the information is important. As a brief terminological reminder, "WGI" stands for within-group interval (time between taps *within* a group), and "BGI" is between-group interval (the time between taps belonging to *separate* groups).

B.1 Data

This section includes details of the extraction of dependent variables from the raw data files, and how some potential problems were resolved. A total of 6 experiments was run in the course of this research, not all of which were reported in the main body of the text. Data analysis involved 3 somewhat different conditions, each of which has somewhat different data extraction and analysis requirements.

1. Isochronous (regular) tapping (e.g., Experiment 1): In this case, it's difficult to define an "error" measure (there are no gaps into which a mistake can be inserted), but it is possible to define fairly sophisticated timing measures (e.g.,

autocorrelation).

2. Two-group patterns: The 4-2 pattern in the bulk of the experiments contains two separate groups, a 4 group and a 2 group. This complicates some forms of analyses quite a bit: for instance, at some level it might be necessary to look at the WGI in each of the groups separately. Error determination is complicated: as noted earlier, a 3 group is ambiguous in this particular case; it could be either a 4 group with a deletion, or a 2 group with an insertion. These problems will be discussed and resolved below.

3. Single group patterns: Although no single group data is reported in this thesis, single group patterns of 3, 4, and 5 taps were used in some other experiments. Analysis here is more straightforward than with two-group patterns. Unfortunately, the 3 group gives much less impairment than the 4-2 group. Interestingly, experimental data left out of this dissertation for various reasons suggest that longer single groups (e.g., groups of 5) show a higher degree of impairment than groups of 3.

B.1.1 Data Filtering

Before quantitative analysis, all trials were visually inspected, and trials in which a base level of synchronization was not achieved were removed from further analysis (the rationale here is that you want solid synchronization before the disruption is introduced). The criterion for an acceptable trial was that there be correct tapping of the final 2 groups (a complete 4-2 pattern) in the synchronization, where correct tapping was defined strictly in terms of count, not placement on the pacing signal.

For the remaining trials, exceedingly short ITIs (due to a brush of an adjoining key in the process of pushing the target key) were removed from all trials before aanalysis; the criterion was (somewhat arbitrarily) set at 20 ms.

Since a partial group at the end of the trial may be due to termination of the (fixed-time) trial rather than an error, the last group was always removed before further analyses.

B.1.2 Grouping Issues

The grouping criterion was described in Section 4.1.5; it involved a fixed BGI cutoff. This sometimes classified as errors groups that a visual inspection would have classified as correct, because the subjects had slowed down too much. It is possible that this measure should be made sensitive to subjects' time drifts, but that obviously would be much more complicated.

Fortunately, drifts were generally small. If there *were* a large slowdown (e.g., waiting for feedback from each tap), this might tend to overestimate deletions. This is one of the appeals of insertions as a measure, as they are not subject to such artifacts.

Note that such grouping errors may also affect timing measures (e.g., the mean BGI and WGI reported here), since the same drift that will classify slow groups as errors will also classify possible WGIs as BGIs. On the issue of timing measure complexities, note also that the "unit-to-unit" time of Chase et al (1961b) mixes BGIs and WGIs, which creates a somewhat meaningless measurement. Although I do not rely on timing measures, it's worth noting that more precise measures might show more differences, and also that the current data at least hints that there may be differences between DAF effects on BGI and WGI. This might be crucial for general theories of how such grouped patterns get generated, something I did not have a chance to address.

B.1.3 Error Measure Implementation

The error measures are somewhat more complicated than the description in Chapter 4 suggests. This section will provide some more detail on three issues. Note that in the main analyses I have sometimes only discussed the first (4 tap) group of the 4-2 pattern; this is because otherwise the data becomes somewhat overwhelming. The reported data for both Experiments 1 and 3 show that this is in fact where most errors occur, a datum which is of some interest in itself.

Phase Shifting

The total error measure reported for the 4-2 group in Experiments 1 and 3 was defined as "any group which did not contain either 4 taps or 2 taps" (Criterion A). Clearly, a trial in which a subject continually tapped groups of 4 should be judged as containing errors, but they would not be reported by this measure. On the other hand, requiring the groups to strictly match a sequence of "424242..." (Criterion B) would be too strict, as a single error might cause every succeeding group to be counted as an error. An early form of error measure (Criterion C) addressed this by allowing for a phase shift, that is, following an error group (which *was* defined by sequence, that is, a 4 group following a 4 group would be an error), the following group was allowed to be either a 4 or a 2, so that the phase could be "reset" if necessary (this will give a higher error measure than Criterion A). As it turned out, Criterion A and Criterion C gave measures which were very close. For example, in Experiment 3, Criterion A applied to the 250p condition gave an overall error count of 2.73, compared with Criterion C's measure of 2.70 (a difference of 1%). In the 400p condition, Criterion A gave an overall error count of 2.10, compared with Criterion C's measure of 2.03 (a difference of 3%). Given these small magnitudes, adaptation of the simpler measure seemed justified.

Ambiguity

Another issue is the classification of 3 groups, which are ambiguous. The procedure used here, in which a 3 group was classifed as an insertion only if it followed a long group (4 or more taps) will potentially *underestimate* the true number of insertions on the 2 group. This is of particular relevance to the claim made that insertions on the 4 group occur much more often than insertions on the 2 group (something which did not relate to the main issues in this thesis, but which is of interest for other reasons).

Again, actual inspection of the data suggests that this issue is more theoretical than actual, and the conclusion is not compromised. E.g., looking again at the Experiment 3 data, mean total single insertions in the 250p condition were calculated as 1.90, made up of 1.32 4-group insertions and .58 2-group insertions. Totalling *all* the 3 groups in the trial (thus, presumably *overestimating* 2-group insertions) gave a total of .73. For the 400p condition, total single insertions were calculated as 1.08, made up of .92 4-group insertions and .16 2-group insertions. Totalling *all* the 3 groups in the trial gave a total of .30. Thus, although the proportions change if the overestimate is used, the 4 group is still the predominant source of insertion errors.

Note that since the miscellaneous error measure is *defined* as total errors minus single insertion errors, 3 groups which are not unambiguous single insertions by this criterion are included in the miscellaneous count.

Normalized Counts

I report *total* error counts throughout, rather than measures normalized to time. Since all trials in an experiment had their length defined in number of patterns (rather than elapsed time), and since (except for Experiment 1) this was also matched across experiments, the comparison of such raw counts across conditions, rates and experiments is valid. Direct comparison across different patterns is a more dubious comparison, but not one of importance to the data presented here.

Such measure could perhaps be normalized, as has sometimes been done in speech (e.g., a "correct word" rate, which combines rate and time). However, such measures tend to rely on the fact that speech research usually has a fixed text, so that the number of possible correct items is fixed in advance.

B.2 Presynchronization and the SAF condition

As briefly noted in Chapter 4, there is unfortunately a potential problem with the timing data in the 0 ms delay (SAF) condition; there is a possible speedup artifact (and since the the Experiment 1 data the SAF condition is involved in the main differences, the use of the timing data in general is weak). The problem, briefly, is as follows: When subjects synchronize to a regular pacing sequence, it is robustly established that they tend to tap a small amount (e.g., 40 ms) ahead of the sounded tone; I will refer to this as "presynchronization". In the current experiment, subjects synchronized to the pacing tone without hearing themselves, and then are suddenly shifted to a condition without pacing tones, but with feedback to their tones (in the SAF condition, to *immediate* synchronous feedback). The crucial issue is that (due to the 40 ms presynchronization) this leads to a small phase advance in the timing pattern of heard tones. Although the pitch of the tones has changed, and subjects are told that they are supposed to maintain their same tapping rate, it cannot be ruled out that this might have some effect of causing subjects to shift phase, which could lead to a speedup. For instance, Fraisse and Voillaume (1971) performed an experiment in which subjects were required to synchronize with a tone and then were shifted (*without notice*) to synchronous feedback at the same pitch. Subjects often sped up dramatically. Although there are many crucial differences between that experiment and this one (e.g., their subjects weren't told that the tones would shift from pacing to feedback, and synchronization was defined as task throughout the trial), it is still a possible concern.

One way of establishing that this isn't a confound in the current data would be to demonstrate that the SAF condition has the same timing characteristics as a no-auditory feedback condition, since the auditory phase shift is not present in the latter case. There is certainly no difference in errors, as Experiment 1 shows; and Experiment 1 also shows little timing difference. However, an unreported replication in Experiment 4 *did* show a timing difference in these two cases, so the issue is unresolved.

There are a number of possible workarounds. The choice of no feedback during the synchronization phase was somewhat arbitrary, and experiments could be rerun in which subjects got SAF in this part of the experiment. Or, as suggested by Bruno Repp (personal communication), it would be possible to *delay* the SAF in the continuation phase by a small amount (e.g., 40 ms), to avoid the phase shift, though this has complications of its own and would probably need to be done on a per-subject basis, since the amount of presynchronization varies across subjects. But for now, the (small) effects in the timing data are simply not seriously analyzed.