Inequality in identification of direction of frequency change (up vs. down) for rapid frequency modulated sweeps

Michael Gordon
Biology Department, University of Maryland, College Park MD 20742
mgord@wam.umd.edu

David Poeppel
Cognitive Neuroscience of Language Lab - Linguistics Department, Biology Department, University of Maryland, College Park MD 20742
dpoeppel@deans.umd.edu

Abstract: The abilities of human subjects to identify direction (up vs. down) of frequency modulation (FM) of individual tone sweeps at various rates of FM are examined, in particular, how fast FM sweeps can be without impairing a subject’s ability to accurately identify them as upward or downward. This ability may be relevant to the auditory encoding of rapid formant transitions, important perceptual cues in speech sounds. In a single-trial 2AFC task, subjects identified randomly presented FM sweeps by pressing one of two labeled keys (up or down). Subjects were significantly better at identifying upward sweeps than downward ones at rapid FM rates (6.2 oct./sec. – 25.0 oct/sec., parameterized as stimulus duration at constant bandwidth).

© 2000 Acoustical Society of America
PACS numbers: 43.66.Mk
Date received: 14 April 2001 date accepted: 4 October 2001

1. Introduction

Formant transitions are components of speech sounds that change rapidly in frequency (over 10s of ms) and are known to be important cues for speech perception. Direction and rate of frequency modulation (FM) are among the features of formant transitions that influence speech perception. A well studied category of speech sound for which formant transitions are important cues are the stop consonants (b, d, g, p, t, k). For instance, the direction of the F2 transition is a primary cue for distinguishing /ba/ from /da/ (Delattre et al., 1955).

Unlike with isolated FM sweeps, listeners do not consciously hear the rising or falling tonal quality of formant transitions while listening to speech. It is likely that the neurally encoded acoustic information of formant transitions is integrated with that of other components of the speech sounds, and this integration leads to the percept of a speech sound, as demonstrated by the phenomenon of duplex perception (Liberman et al., 1981). Whether this encoded signal contributes to the perception of an FM “chirp” or to a speech sound depends on the context in which it is found (in isolation or with other speech sound components). Nevertheless, since the spectro-temporal acoustic structure of formant transitions is known to be a strong cue in speech perception, the human auditory system is presumably able to initially encode this information in a reliable way.

Because of the rapidity of the frequency change in formant transitions, they present a spectro-temporal encoding challenge for the auditory system. In fact, there is some evidence that deficits in temporal auditory processing, including processing of FM, are associated with phonological impairments in certain individuals (Tallal, 1980; Talcott et al., 2000).

To deal with the spectro-temporal encoding challenge of rapid formant transitions, the human auditory system may rely on a class of central auditory neurons specialized for...
processing FMs. Such neurons, widely found in animal models, are selectively tuned to features of FM such as direction and rate of change. Evidence that such specialized auditory channels exist in humans comes largely from selective adaptation studies (Tansley and Suffield, 1983). It is possible that deficits in temporal auditory processing of FMs may reflect a defect in this subclass of central auditory neuron (Stefanatos, 1989).

In this study, we tested the spectro-temporal processing limits of the human auditory system with regard to identifying the direction of FM in brief isolated tone sweeps: how fast (short) can FM sweeps be and still be identified correctly by human listeners as upwardly or downwardly sweeping tones? Although this task differs markedly from that of identifying speech sounds, it tests spectro-temporal processing capabilities of the auditory system that are relevant to speech sound processing. We observed an asymmetry in subjects’ ability to identify FM direction between upward and downward sweeps. Specifically, subjects were better at identifying upward sweeps when the FM rate was rapid (short duration).

2. Materials and methods

The stimuli were linearly frequency-modulated tone sweeps (glides), generated with Matlab (Mathworks, Natick MA). Both upward and downward FM sweeps of 10 different rates of frequency change (FM rate) were presented in each of the 3 frequency ranges tested (0.6-0.9 kHz, 1-1.5 kHz, 2-3 kHz). The bandwidth of all FM stimuli was kept constant at half an octave to approximate the bandwidth of formant transitions. The three frequency ranges were selected because they span typical values for F1, F2, and F3 values in speech. For each frequency range tested, the stimulus set consisted of 20 different FM sweeps (2 directions x 10 FM rates). The different FM rates were created by varying the duration of the stimuli (bandwidth kept constant). The following 10 durations were used: 5, 10, 20, 30, 40, 50, 80, 160, 320, 640 ms. These durations gave the following FM rates: 100.0, 50.0, 25.0, 16.7, 12.5, 10.0, 6.2, 3.1, 1.6, 0.8 oct./sec. The relative intensity of the stimuli was set to compensate for the duration-loudness trade off to make all stimuli of roughly equal loudness; the shortest FM stimulus (5 ms) was four times more intense than the longest stimulus (640 ms; Port, 1963). This was done because we were interested in studying the effect of FM rate rather than the effect of stimulus loudness on subjects’ ability to judge FM direction. All FM stimuli had a linear rise-fall time of 2 ms to alleviate spectral splatter.

Subjects were normal-hearing undergraduates from the University of Maryland with no history of hearing or neurological problems. The stimuli were presented binaurally through Sennheiser HD 520 headphones by a Macintosh G4. Presentation of stimuli and collection of responses was controlled by the PsyScope experimental platform (Cohen et al., 1993). Stimuli were presented at comfortable supra-threshold levels ranging from about 65-77 dB SPL (measured with a Bruel and Kjaer Type 2232 sound-level meter).

Each of the 3 frequency ranges was tested in a separate experiment. For each experiment, about thirty subjects were run; one set of subjects was run at the mid-frequency range and a second set of subjects was run at both the low and high frequency ranges. In a single-trial 2AFC task, subjects had to identify the direction of FM by pressing one of two labeled keys on a computer keyboard. Before each experiment, subjects were given a 30-second practice session (19 trials) to familiarize them with the stimuli. Each experiment consisted of 400 trials (20 repetitions of 20 FM stimuli), presented in pseudo-random order. The inter-trial interval was varied between 750, 1000, 1250, 1500, and 1750 ms. Subjects were given a short rest period after 200 trials.

3. Results

There was a marked difference between subjects’ abilities to identify upward and downward FM sweeps of rapid FM rate (short duration) for all three frequency ranges tested. Subjects were more accurate at identifying upward FM sweeps than downward FM sweeps when the
FM rates were greater than 3.1 oct./sec. (durations shorter than 160 ms). This effect was strongest in the case of high and mid-frequency ranges.

In the case of the high frequency range (Fig. 1), when FM rates were at the lowest values used (0.8, 1.6, 3.1 oct./sec.), there was no difference in FM direction identification performance between upward and downward FM sweeps. Performance was at ceiling for these slowest FM rates at about 95% identification accuracy for both directions of FM. However, at faster FM rates of 6.2 oct./sec. and greater, subjects were considerably more accurate at identifying upward sweeps compared to downward ones. When sweep rate was increased to 25 oct./sec. and greater, identification ability again converged roughly at chance, indicating that, for both directions of FM these shortest stimuli were too rapid for humans to be able to identify FM direction. Although FM direction identification performance increased with decreasing FM rate (increasing FM duration) for both upward and downward sweeps, there was a much sharper increase in performance with decreasing FM rate for upward sweeps. At the rate of 16.7 oct./sec., subjects were already at ~90% correct in identifying upward sweeps (Fig. 1, left). In contrast, downward sweeps needed to be as slow as 3.1 oct./sec. for subjects to correctly identify them with an accuracy of 90%. An analysis of variance testing direction and duration/rate confirms a main effect of duration/rate [F(9, 23)=183.6, p<0.0001], a main effect of direction [F(1, 23)=245.4, p<0.0001], and, crucially, a significant direction x duration interaction [F(9, 23)=21.3 p<0.0001].

Fig. 1. Results for high frequency range (2-3 kHz), n = 32, error bars denote S.E.M.

The inequality in FM direction identification between upward and downward sweeps was not reflected in the reaction time data. (These reaction time values are relative to the onset of stimuli.) There was not much difference between upward and downward sweeps with respect to reaction times (Fig. 1, right panel). For both FM directions, there was a gradual decrease in reaction time as the FM rate decreased from 100 to about 16.7 oct./sec., after which further decreases in rate affected only a very slight further decrease in reaction time. At the slowest rate used (0.8 oct./sec.) there was a reversal in the trend, and the reaction time increased. The analysis of variance revealed a main effect of duration/rate [F(9, 23)=26.7, p<0.0001], a significant direction x duration interaction [F(9, 23)=5.3, p<0.0001], but no main effect of direction [F(1, 23)=0.98, p=0.32].

The FM direction identification performance observed for the mid- (1.0-1.5kHz) and low (0.6-0.9kHz) frequency ranges (Figs. 2 and 3, respectively) mirrors the performance for the high frequency range. As in the case of the high frequency range, subjects were better at identifying upward FM sweeps than downward ones when the rates were greater than 3.1 oct./sec. For the mid-frequency range, this was true up to rates as fast as 25 oct./sec., just as for the high frequency range. For the low frequency range, the “upward FM advantage” persisted up to rates as fast as 16.7 oct./sec. The middle range showed significant main effects of duration/rate and direction as well as a significant interaction [F(9, 21)=150.5, p<0.0001]; [F(1, 21)=143.1, p<0.0001]; [F(9, 21)=37.3, p<0.0001], respectively. The low
range was associated with roughly parallel results \[F(9, 23)=182.3, p<0.0001\]; \[F(1, 23)=5.85, p=0.015\]; \[F(9, 23)=23.7, p<0.0001\].

As was the case for the high frequency range, the reaction time data for the mid frequency range showed no significant difference between upward and downward sweeps \[F(1, 21)=1.52, p=0.22\]. In the low range, the ANOVA revealed a main effect of direction \[F(1, 23)=9.32, p=0.002\]. Post-hoc tests revealed that the effect is due to the 3 most rapid FM rates. The reaction time data for mid and low frequency ranges do illustrate, better than the high frequency range data, how reaction times were relatively constant across FM rates ranging from 1.6 to 16.7 oct./sec.

The performance data seen for the mid- and low frequency ranges differed from the high frequency range performance data in that there was a crossing at the fast end of the rate continuum (at about 50 oct./sec. for mid-frequency range, at about 25 oct./sec. for the low frequency range) of performance curves for up and down sweeps. At the very fastest rates, subjects actually showed a slightly greater ability to identify downward FM sweeps than upward ones. It should be noted that these values are nevertheless near chance, so interpreting the meanings of these crossings is difficult. For the fastest rate FM sweeps used (100 oct./sec.), it was extremely difficult for subjects to identify any sense of frequency change in the stimuli at all.

4. Discussion

Although we chose to discuss our results primarily in terms of FM rate, it must be noted that the results cannot be attributed solely to effects of FM rate, because FM sweep duration co-varied with the parameter of FM rate. Our results should therefore be interpreted in the context of FM sweeps that varied in FM rate and duration and had a constant bandwidth of 0.5 oct. in frequency ranges relevant to speech sounds.
Humans are better at identifying rapid upward FM sweeps than downward ones in the range of FM sweep rates from 6.2 oct./sec. to 25.0 oct/sec. FM sweeps that are faster than this range are perhaps beyond the spectro-temporal resolution of the auditory system, and hence both upward and downward sweeps are identified at near chance performance levels for such rapid sweeps. FM sweeps that are slower than this range are also identified with equal performance, perhaps because the task becomes very easy at very slow sweep rates. Hence performance is at a peak level (about 95% correct identification) for both upward and downward sweeps at such slow FM rates. Our results, which were obtained with an FM sweep direction identification task, are consistent with a previous FM detection study; Collins and Cullen (1978) revealed lower thresholds in human subjects for detecting rapid upward FM sweeps than for downward ones, indicating an increased sensitivity to upward sweeps.

The perceptual advantage that upward sweeps have over downward sweeps is reflected in physiological measures (evoked potentials) of auditory processing at multiple levels of the auditory system. Compound action potentials, which represent synchronous VIIIth nerve activity, were measured in the guinea pig in response to exponentially rising and falling rapid FM sweeps. Upward sweeps elicited a greater amplitude response than downward ones (Shore and Nuttall, 1985). In another study, exponentially rising and falling sweeps were presented to human subjects while auditory brainstem responses were measured. Again, upward sweeps elicited a larger response than downward ones (Dau et al., 2000). In a study which focused on cortical auditory responses to linear FMs, there was a greater amplitude response to upward sweeps than to downward ones (Maiste and Picton, 1989).

As discussed by Shore and Nuttall (1985) and more recently by Dau et al. (2000), the difference in evoked auditory responses to upward and downward sweeps most likely originates at the level of the basilar membrane. Because frequencies are spatially represented along the length of the basilar membrane, there are frequency-dependent temporal delays imposed by the traveling wave, with high frequencies (basal end) having a shorter latency than low frequencies (apical end). Thus, for a broadband stimulus like a click, in which spectral energy is presented simultaneously across frequencies, high frequency regions of the basilar membrane will be excited earlier than the low frequency regions. This results in a desynchronization in the firing of auditory neurons across frequency channels. For downward FM sweeps, this asynchrony is exaggerated because high frequencies are presented before low frequencies, which compounds the effect of the delay imposed by the traveling wave. Upward sweeps, on the other hand, compensate for the temporal delays imposed by the traveling wave and result in a more synchronous firing of neurons across frequency channels. The increased synchrony of neuronal firing with upward sweeps leads to greater amplitude evoked potentials for these stimuli.

The exponential sweeps of large frequency excursion used by Shore and Nuttall (1985) and Dau et al. (2000) were specifically designed to match the delay line characteristics of the basilar membrane. This was done to optimize the ability of upward sweeps to compensate for the temporal delays imposed by the traveling wave. However, the difference in reactions of the basilar membrane in response to upward vs. downward FM sweeps may be large enough to affect perception even in the case of less optimized stimuli, such as linear FM sweeps of shorter frequency excursion (as in this experiment), as long as they are of the appropriate FM sweep rate.

The approximate duration at which we found upward FM identification to become severely compromised (20-30 ms) matches fairly well the threshold for perception of temporal order for pairs of tones (Hirsh, 1959). Indeed, a recent study suggests that both temporal order judgments (near threshold) and auditory processing of rapid FM sounds may rely on a common neuronal mechanism – frequency change detectors, also known as “FM processing” neurons (Okada and Kashino, 2001). This study examined the effect of exposing subjects to FM adapting tones on their ability to correctly identify temporal order of pairs of tones (low and high frequency). Upward FM sweeps were the adapters in one condition, and
downward sweeps were the adapters in a second condition. The temporal order judgment of the subjects was affected by the adapting FM tones in a FM direction – specific way. The point of subjective simultaneity along the tone-onset-delay axis was shifted in one direction (low tone preceding high tone) by upward FM adapters and in the opposite direction (high tone preceding low tone) by downward FM adapters.

The duration of ~25 ms may constitute a generalized temporal threshold in auditory perception. A recent study examined electrophysiological and psychophysical responses to click trains with interclick intervals ranging from 1 to 1000 ms. A sharp transition in both perception and physiology occurred as the interclick interval went from above 25 ms to below 25 ms (Boemio, Poeppel and Hodos, 2000).

The ability to identify rapid upward FM sweeps may be limited by the spectro-temporal integration limits of specialized auditory neurons, whereas the ability to identify rapid downward FM sweeps may be *further* limited by temporal delays imposed by the traveling wave in the cochlea. Given the inequality between rapid upward and downward FM sweeps with regard to perception and physiological responses, one might ask how this relates to the auditory encoding of formant transitions and the frequency of occurrence of formant transitions in a language. There is no a priori reason to assume that upward formant transitions are more salient than downward transitions. Similarly, there is no reason to assume that speech sounds that incorporate upward transitions are more frequent than those with downward transitions. Given the asymmetries in FM processing that have now been documented behaviorally and electrophysiologically, the interaction between FM processing and speech sound processing will have to be further explored.

**Acknowledgment**

This work was supported by a postdoctoral traineeship to Michael Gordon from the NIDCD Training Program in Comparative and Evolutionary Biology of Hearing (CEBH) at the University of Maryland at College Park.

**References**


