Question 1

The control of the fundamental frequency of a speech signal can be important in both singing and in speech production. Our ability to control the production of laryngeal vibration is, to a great extent, dependent on our ability to extract that information perceptually from a complex speech signal.

Discuss how fundamental frequency of such complex signals is encoded at various levels of the auditory system. Relate that discussion to our ability to perceive changes in fundamental frequency in psychophysical experiments with relatively simple stimuli as well as with speech.

INTRODUCTION

Since the function of singing and speech is to provide a vehicle for communicating thoughts, the auditory signal could be one of the most important factors used in controlling the speech motor system. The speech motor control system has been shown to have certain aspects dependent on sensory neural feed back (Gracco and Abbs, 1985, Shaiman, 1989). Specific studies have displayed that when phonating subjects hear their voice pitch feedback shifted, they respond with a change in voice fundament frequency output (Burnett *et al.*, 1997, Burnett *et al.*, 1998). As stated in the comprehensive exam question, "the control of the fundamental frequency of a speech signal can be important in both singing and in speech production. Our ability to control the production of laryngeal vibration is, to a great extent, dependent on our ability to extract that information perceptually from a complex speech signal."

Perceptually extracting the fundamental frequency (Fo) from a complex speech signal has been attributed to the central processing of the brain after obtaining signal information encoded via the inner ear and auditory nerve (Houstma and Goldstein, 1972). Investigations of the auditory system encoding were reported in 1954 with the complex encoding studies beginning in the 1960's. Since then, many studies have been done, especially with the conception of the cochlear implant. This paper reviewed the method of encoding Fo at various levels of the auditory system for simple and complex, speech-like, signals. That review was then related to the human ability to perceive changes in fundamental frequency (ÄFo) in psycho-physical experiments with simple stimuli and with speech.

I. AUDITORY SYSTEM AND FUNDAMENTAL FREQUENCY

A. Encoding of Acoustic Stimuli

The Auditory system encodes frequency information from an acoustic stimulus as both temporal and spectral cues. This system consists primarily of three parts: the inner ear (cochlea), the auditory (VIIIth) nerve, and the auditory regions of the brain (auditory cortex). The spectral resolving characteristics of the system begin at, and are primarily due to the frequency selectivity of, the cochlea, while the auditory nerve acts mainly as the communication line allowing acoustic information transmitted to the auditory cortex for higher level processing. There appears to be two methods for encoding pitch information; (1) tonotopically preserved vibration place information in the inner ear and (2) neural rates of action potential, or "phase-locking", in the auditory nerve[1].

1. Vibration Place

Acoustic sound gets to the inner ear through a series of impedance matching devices in the outer and middle ear. Once to the inner ear, the sound impedes on the basilar membrane that is inside the cochlea. The basilar membrane varies in material properties from cochlea base to apex. This results in the basilar membrane varying spatially in resonance frequency, with high frequencies associated with the base and low frequencies with the apex (Rhode, 1971). A stimulus of a certain frequency will result in a traveling wave on the basilar membrane, with the resonance area corresponding to the stimulus frequency exhibiting the largest displacement amplitude. With the help of some active properties (Hudspeth, 1997), the basilar membrane can be thought of as a bank of sharply tuned band pass filters.

Along the basilar membrane are inner and outer hair cells. Hair cells are mechano-chemical transducers. They are responsible both for the transduction of basilar membrane vibratory motion to the VIIIth nerve and the active properties of the basilar membrane. For the

---

[1] Note that these two mechanisms depend on many aspects of the stimulus, including level. In this review of the auditory system, all stimuli were assumed to be well above perceivable threshold yet not near suprathreshold.

most part, reception hair cells preserve the sharp tuning curves of the basilar membrane during transduction (Narayan *et al.*, 1998).

The auditory nerve, because of the innervation of the nerve to the hair cells, preserves the tonotopic organization of the basilar membrane, by each nerve representing an area of the basilar membrane. Thus, spectral information is passed tonotopically on the way to, through the auditory nerve, and into the auditory cortex. The tonotopic organization of the VIIIth nerve is not the only property that caries frequency information to the higher processes. Pitch information can also be encoded by the timing of the nerve firings or action potentials along a nerve.

2. “Phase-locking” of Nerve Firings

Javel (1980) discussed the use of action potential timing on the VIIIth nerve to encode pitch information. He explained that through the transduction of the basilar membrane's vibratory motion into auditory nerve action potentials, action potential firing rates are "phase-locked" to the vibration. This phase-locking is believed to happen at the synapse. Phase-locking is defined as when neural spikes on the auditory nerve occur within a time centered distribution of integer multiples of the vibration period.

Phase locking, as shown by Javel, plays a role in decoding a stimulus' temporal characteristics. A limitation of phase locking happens at higher frequencies. At higher frequency, the time jitter responsible for the distribution plays a larger role as stimulus period becomes smaller. Note that synapses and the auditory nerve also have tuning ability on the incoming information, yet the original tuning of the basilar membrane is mostly preserved.

Spectral and temporal information traveling on the auditory nerve, which is tonotopically arranged, continues to the auditory cortex. In the auditory cortex many cells preserve the tonotopical arrangement. A variety of cells at this level look for different cues in the incoming information; e.g., frequency modulation, onset, offset, etc. (Pfeiffer, 1966, Pickles, 205-214).

B. Pitch in the Auditory System

Pure tone information can be encoded using the mechanisms described above. When complex signals vibrate the basilar membrane, the mechanics are not straightforward because the basilar membrane is nonlinear[2]. For complex harmonics such as speech, the basilar membrane can be thought of as a nonlinear frequency analyzer with neural outputs, where multiple components, in a first approximation, superimpose on the membrane.

Some limitations of the encoding mechanisms need to be mentioned for the case of multiple frequency components. First, within the frequency selectivity of the basilar membrane are critical-bandwidths in which two components within that bandwidth may not be resolved spectrally (Fletcher, 1940). Second, phase-locking does not resolve temporal information at higher frequencies, so higher frequency components of a complex signal may not be resolvable.

Complex signal pitch information can be conveyed by either the spectral (tonotopical) and/or temporal (phase-lock) information (Dorman, 1994, Kiang, 1980, Houtsma and Smuzynski, 1990, Ohgushi, 1978, Simmons, 1992, Turner et al., 1983, Winkler, et al, 19XX). For example, Delgutte and Kiang (1984) showed that period histograms from measured discharge patterns of auditory nerve fibers, for two formant vowels, usually had the largest temporal component of either; the fundamental frequency, one of the formant frequencies, or the characteristic frequency (CF) of the nerve. The position of a fiber's CF relative to the fundamental frequency or formant frequencies appeared to determine which component dominated the response. These experiments show that the fundamental frequency information of a complex signal can be encoded by both rate

---

[2] The nonlinearity of the basilar membrane creates distortion products that can be perceived but they have limited effect on pitch perception

and place in the auditory nerve[3].

Houstma and Goldstein (1972) believed that pitch could be elicited by both temporal and spectral coding. Yet, they concluded that pitch extraction of complex harmonic tones had central processing origins, especially when components were unresolved or the fundamental was missing. They showed this with a simple dichodic situation where subjects had a 1000-Hz tone presented to the left ear and a 1200-Hz tone presented to the right ear. Subjects reported perceiving a pitch of 200 Hz somewhere inside the head that could be matched or recognized.

Whitfield (1980) showed that for cats, pitch perception of complex stimuli happens only in the auditory cortex. Trained cats, without their cortex, responded only to individual frequencies of some complex stimuli. These cats were unable to detect pitch in complex stimuli, which stimuli had elicited pitch response before the cortex was removed. Others have also given evidence of the pitch originating from central processing (Hall, 1986).

Some debate continues whether the pitch extraction process uses the temporal and/or spectral information. Pitch extraction from purely temporal cues is less salient than that of spectral cues (McKay & Carlyon, 1999). Also, questions have been raised whether the mean rate cue in temporal coding by itself supports musical interval recognition and production (Carlyon, 1997). On the other hand, hearing-impaired listeners seem to use the temporal information better than spacial during pitch discrimination tasks (Arehart, 1994). Strong evidence that both temporal and spectral encoding of fundamental frequency, together or individually, can be used for pitch encoding for extraction (Dorman, 1994).

Speech and singing signals are harmonic complex waveforms. To help with vocal control,

---

[3] An extensive review of speech representation in the auditory system can be found in Sach, Winslow, and Blackburns' chapter entitled, "Representation of speech in the Auditory Periphery."

pitch information from harmonic signals can be extracted. Fundamental frequency information is carried through the auditory nerve from the basilar membrane to the higher process of the auditory system. This information is encoded both by temporal (with some limitations at higher frequencies) and spectral (tonotopic or place, but with critical bandwidth limitations) cues[4]. Processes in the auditory cortex then extract the fundamental frequency from rate and place information.

II. OUR ABILITY TO PERCEIVE CHANGES IN PITCH

A. Introduction to the Perception of Pitch

Our perception of both speech and singing pitch is directly to our sensitivity to changes in the pitch of complex signals. It was discussed above how fundamental frequency of speech and singing signals may be encoded at various levels of the auditory system. In this section, the human ability to perceive changes in the fundamental frequency from relatively simple stimuli and speech type stimuli was the topic. Faulkner (1985) found that pitch discrimination of complex signals depends on factors affecting the discriminability of simple signals. In discussing perceived changes in fundamental frequency, psychophysical experiments outlining our ability to perceive changes in simple tones and pitch from harmonic tones were examined. This was done first for simple stimuli and then speech. These experiments were then related to the earlier review on pitch encoding.

1. A note on Pitch Discrimination

Fletcher (1940) showed the existence of a critical bandwidth (about 10% difference in Hz) in which multiple simultaneous tones within that bandwidth cannot be. On the other hand,

---

[4] These limitations categorize stimuli as “resolved” and “unresolved” components. The perception of the pitch is better for “resolved” components.

sequential differentiation of pitch can be done to 1/10 of 1% in Hz, despite critical band theory. In singing or speech, pitch can be thought of as sequential. This allows us to be able to use our higher sensitivity to sequential pitch differences. Yet, how does the relatively continuous nature of pitch in phonation compare with our ability to differentiate sequential pitches?

Srulovicz and Goldstein (1983) discussed pitch encoding and presented a curve of just-noticeable-difference in frequency. This curve shows that for tone discrimination, we are most sensitive between 400-3000 Hz, where thresholds increased for lower and higher frequencies. The low frequency increase was contributed to not enough action potentials on the nerves to extract pitch adequately, while the high frequency increase was contributed to the loss of resolution in phase-locking.

2. A note on temporal acuity

Previously it was discussed how Fo is encoded in the auditory system. In the last section, it was mentioned that perception change in pitch depends on sequential differentiation. Because of this dependence, a few notes on the temporal acuity of the auditory system were mentioned. Two effects that may play a part in perception of Fo change in time are (1) forward masking, which is an increase in the perception threshold of some stimuli following some previous stimuli of length up to 200 ms, and (2) fatigue, which is a similar masking effect but for previous stimuli of longer than 200 ms. When perceiving cyclic changes in fundamental frequency (ÄFo), these effects and may have some affect. Relkin and Turner (1988) showed that forward masking was most likely a central auditory system effect and not a product of the peripheral system. Perception of ÄFo, which involves the central system, may be dependent on forward masking.

B. Perceptually Discriminating Pitch Change

In speech, pitch is variable in time and can be manifested both as a large change, like a glide, or a slight neurological perturbation or modulation (cyclic), whether desired or pathological (Titze, et al., 1987, Titze, 1991). To address perceived changing pitch in voicing, psychophysical experiments addressing two types of pitch changes have been reviewed: frequency modulated (FM), and non-FM (long term) variations. This was done for both simple stimuli and speech-like stimuli.

1. Simple Stimuli

a. Frequency Modulation Perception

Many studies have been completed involving temporal cue pitch tracing through frequency modulation means (Sek and Moore, 1995, Houstma and Smurzynski, 1990). The most simple stimuli that could be frequency modulated is a sinusoid. One example is Demany and Semal (1988) who looked for detection thresholds for sinusoidal frequency modulation. Their stimuli were several sinusoidal carriers, 250-6000 Hz, modulated with a variety of frequencies, 1-64 Hz. Modulation detection thresholds were measured at ~3% at the lowest carrier and modulation frequencies or, in other words, thresholds were 3% modulation of the lowest carrier frequency.

This is similar to what Zwicker (1952) found. Zwicker showed that the lowest threshold was for 1-20 Hz modulation and decreased for more modulation. So according to these two examples, thresholds increased as carrier and/or modulation frequency increased.

Sinusoidal frequency modulation would be encoded in the auditory system by one of the two methods mentioned above. The auditory cortex would need to notice most of the action potentials oscillating in adjacent nerves in the tonotopic arrangement for a spectral cue, or it would need to notice a temporal variation in the rate of the action potentials along a specific fiber. For modulations less than the threshold, forward masking and fatigue may be the underlying effect.

Dimany and Semal (1998) found an interesting event of particular interest to fundamental frequency control. For low modulation frequencies, FM detection is mediated by a pitch sampling process using a temporal window of about 25 ms. This may have implications on the speed that modulation detection takes place

b. Non-FM Perception

Dooley and Moore (1988) measured thresholds for the detection of differences in duration for sinusoids whose frequency was swept up or down by 100 Hz, centered at 2000 Hz. They found that the thresholds for gliding-tone conditions were between 24 and 34 ms with a slight dependence on glide direction. This is comparable to the FM pitch detection sampling temporal window of 25 ms.

A 25-ms sampling process is not limited by the periphery of the auditory system, which can encode information at a much higher rate than this, so it must have something to do with the pitch extraction. For low modulation frequencies, where speech Fo variations happen, the fundamental frequency detection is sampled, implying that the central processing required to extract speech may be organized to notice variations slower than 25 ms. This sampling window is enough to include pertinent varying pitch information in speech and singing. Most neurological perturbations of fundamental frequency in speech occur on average between 3-12 Hz.

Madden and Fire (1997) also studied the detection and discrimination of frequency glides with similar results as above. They accounted for their results by using a place mechanism of frequency coding based on "the detection of changes in the excitation pattern by a signal." Relating this back to the previous section on fundamental frequency encoding in the auditory system, this could be a mechanism that tracts the spectral information across the tonotopical arrangement. Possibly one cell such as Pfeiffer (1966) found in the auditory cortex would look for these changes or variation in place as a function of time.

2. Speech Stimuli

a. Frequency Modulation Perception

Shackleton and Carlyon (1994) investigated the discriminability of frequency modulated harmonic complexes containing harmonics of specific frequency bands creating both a resolved and an unresolved harmonic stimulus set. They measured the frequency-modulated threshold as about 3ÄF_0/F_0 for carrier frequencies of 62.5, 125 and 250 Hz. The threshold increases for increased modulation and decreases with increased carrier fundamental frequency. The threshold also increased if the frequency region, from which the harmonic components were taken, was raised. In other words, higher frequency components did not contribute as well to FM discrimination as lower components. Note that 3ÄF_0/F_0 results in a factor of 3 greater thresholds than the simple tone FM perception.

The effects, larger modulations and carrier Fo, can be described in terms of the previous discussion. Threshold dependence on modulation may be related to (1) critical band theory, which is when the modulations exceed a critical band thus increasing perceptibility or becomes resolvable, (2) the forward masking effect, which is when a certain tonotopical area is still under previous stimulus masking influence when the stimulus returns depending on the modulation frequency, or (3) not enough action potentials at low frequencies (remember the just-noticeable-difference for low frequency mention above). The effect of a frequency region of components on the threshold, where higher frequency components did less for the threshold, can be related to the limitations of the phase-locked action potentials, which loses resolution at high frequencies.

This can be summarized as, the increase in the threshold occurred when harmonics became unresolved, implying that fundamental frequency change perception is better when the periphery can resolve the components of a series, both temporally and spatially. Therefore, as long as phase-locking enables the encoding of temporal cues and multiple frequency components are not within a critical band, discrimination thresholds are low. Nevertheless, when frequency components become unresolved for the auditory system periphery, discrimination thresholds increase.

2. Non-FM Perception

Gagne and Zurek (1988) measured the just-noticeable change in resonance frequency of a filter set that was to be similar to the characteristics of a speech sound. Filters were resonated with different stimuli: a periodic square wave (comparable to the glottal spectrum), with a smooth change in fundamental frequency (50 Hz in 100 ms), or a random white noise. Resonance stimuli were not found to affect the just-noticeable change in resonance frequency. The filter banks used in their study had a range of resonance frequency (F_{res}) form 300-2000 Hz and quality factors (Q or selectivity) from 1-36. Measured just-noticeable change in resonance frequencies perceived by subjects can be summarized as $\mathbf{0.079\ F_{res}/\sqrt{Q}}$.

Gange and Zurek found that just-noticeable change in resonance frequency increases with the frequency and bandwidth of the resonance. Notice that the just-noticeable change decreases for increased Q that is a wider filter. This is similar to widening the critical bandwidth, which is similar to making frequency components unresolvable spatially.

Just as an example of where we would use our discrimination ability of change in fundamental frequency, this next experiment was reviewed. Hillenbrand and Houde (1996) showed that changing fundamental frequency plays a role in perception of intervocalic glottal stops. They found that a dip in the pitch contour is nearly always sufficient to cue the presence of a glottal stop. In these dips fundamental frequency would change in a dipping manner from average Fo by 18% to 60%, and lasting about 75 ms. This change was perceived so to help perception of the stops.

Hillenbrand and Houde also looked for what amount of frequency change was needed to have a synthesized sound perceived as a glottal stop. They measured a group identity function that showed some values of changing fundamental frequency in speech that can be perceived. Chance perception of judged glottal stops for the subjects occurred at a frequency dip of 7% of fundamental frequency with an asymptote of best perception (>95%) at ~12% of Fo. These dips lasted between 230-400 msec. These values are 2 to 3 times the just-noticeable difference values. The difference may be due to the added lingual, or learned, mechanisms similar to categorical perception, where the subjects could have distinguished that something was different but they did not perceive it as a phoneme change until it reach 7%.

CONCLUSION

Change in fundamental frequency can be perceived and measured in both modulation and glide examples. Fundamental frequency perception may be done at the central auditory system, which is dependent some on the ability of the peripheral system to resolve frequency components. The threshold of ÄFo perception may be explained through the effects of the encoding of pitch in the peripheral system, with dependence on temporal acuity.

One encoding system is “phase-locking.” Phase-locking is when neural spikes on the auditory nerve occur within a distribution centered in time at integer multiples of the vibration period. With higher frequencies, the time pulses that occur around the integer multiples and are governed by a distribution, begin to overlap, losing temporal resolution. When the width of the distribution is of comparable size as the period of vibration, phase-locking can no longer be useful as a pitch cue. With the loss of this cue, pitch perception decreases and so does the perception of pitch change.

The other encoding system is the place information of vibration amplitude on the spatially varying frequency selectivity of the basilar membrane. With this system, more modulation decreases the threshold. This may be explained by referring to the limits of spacial resolution power or the critical band. Once the modulation is stimulating frequencies outside the critical band, this encoding system can better resolve the components.

There also may be a dependence on forward masking in the auditory system. For cyclic changes in Fo, the system may be still recovering post stimuli for a particular region when the region is stimulated again by the return of the stimuli. This effect has been shown to be an auditory cortex effect.

REFERENCES

Arehart, K. H., (1994). “Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners,” J. Acoust. Soc. Am. **95**(6):3574-85.

Burnett, T. A., Senner, J. E., and Larson, C. R., (1997). “Voice F0 responses to pitch-shifted auditory feedback: a preliminary study,” J. Voice. **11**(2):202-11, 1997 Jun.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C., (1998). “Voice F0 responses to manipulations in pitch feedback,” J. Acoust. Soc. Am. **103**(6):3153-61.

Carlyon, R. P., (1997)., “The effects of two temporal cues on pitch judgments,” J. Acoust. Soc. Am. **102**, 1097-1105.

Delgutte, B., and Kiang, N. Y., (1984). “Speech coding in the auditory nerve: I. Vowel-like sounds,” J. Acoust. Soc. Am. **75**(3):866-78.

Demany, L., and Semal, C., (1989). “Detection thresholds for sinusoidal frequency modulation,” J. Acoust. Soc. Am. **85**(3), 1295-1301.

Dooley, G. J., and Moore, B. C. J., (1988). “Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change,” J. Acoust. Soc. Am. **84**(4), 1332-1337.

Faulkner, A., (1985). “Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations?,” J. Acoust. Soc. Am. **78**(6), 1993-2004.

Fletcher, H., (1940). “Auditory paterns,” Rev. Modern Physics, **12**, 47-56.

Gagne, J. P., and Zurek, P. M., (1988). “Resonance-frequency discrimination,” J. Acoust. Soc. Am. **83**(6):2293-9.

Gracco, V. L., and Abbs, J. H., (1985). ‘‘Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes,’’ J. Neurophysiol. **54**, 418-432.

Hall, J. W., (1986). “Central adaptation of complex pitch,” J. Acoust. Soc. Am. **80**(6), 1681-4.

Hillenbrand, J.M., and Houde, R. A., (1996), “Role of F0 and amplitude in the perception of intervocalic glottal stops,” J. Acoust. Soc. Am. **99**(6):1182-90, 1996 Dec.

Houstma, A.J.M., and Goldstein, J. L., (1972). “The central origin of the pitch of complex tones: Evidence from musical interval recognition,” J. Acoust. Soc. Am. **51**, 520-9.

Houstma, A. J.M., and Smurzynski, J., (1990). “Pitch identification and discrimination for complex tones with many harmonics,” J. Acoust. Soc. Am. **87**, 304-10.

Kiang, N.Y., (1980). “Processing of speech by the auditory nervous system,” J. Acoust. Soc. Am. **68**(3):830-5.

Madden, J. P, and Fire, K. M., (1997). “Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency,” J. Acoust. Soc. Am. **102**(5), pt. 1, 2920-2924.

McKay, C. M., and Carlyon, R. P., (1999). “Duel temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,” J. Acoust. Soc. Am. **105**, 347-55.

Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A., (1998). “Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae,” Science. **282**(5395):1882-4.

Ohgushi, K. (1978). “On the role of spatial and temporal cues in the perception of the pitch of complex tones,” J. Acoust. Soc. Am. **64**(3), 764-71.

Pfeiffer, R.R. (1966) “Classification of response patterns of spike discharge units in the cochlear nucleus: tone-burst stimulation”, Exp. Brain Res. 1, 220-225.

Pickles, J. O., (1988). *An introduction to the physiology of hearing*. Academic Press, London.

Relkin, E.R. and Turner, C.W. (1988). “A re-examination of forward masking in the eighth nerve”, J. Acoust. Soc. Am., **84**: 584-91.

Shaiman, S. (1989). ‘‘Kinematic and electromyographic responses to perturbation of the jaw,’’ J. Acoust. Soc. Am. **86**, 78-88.

Shackleton, T. M., and Carlyon, R. P. (1994). “The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,” J. Acoust. Soc. Am. **95**, 3529-3540.

Simmons, A. M., Schwartz, J. J., and Ferragamo, M., (1992). “Auditory nerve representation of a complex communication sound in background noise,” J. Acoust. Soc. Am. **91**(5):2831-44.

Srulovicz P., and Goldstein J.L., (1983). “A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,” J. Acoust. Soc. Am. **73**(4):1266-76.

Titze, I. R., Horii, Y., and Scherer, RC., (1987). “Some technical considerations in voice perturbation measurements,” J. Sp. Hear. Res. **30**(2):252-60.

Titze, I. R., (1991). “A model for neurologic sources of aperiodicity in vocal fold vibration,” J. Sp. Hear. Res. **34**:460-472.

Whitfiled, I. C., (1980). “Auditory cortex and the pitch of complex tones,” J. Acoust. Soc. Am. **67**(2), 644-7.