Hearing Research 354 (2017) 16–27

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Stream segregation of concurrent speech and the verbal transformation effect: Influence of fundamental frequency and lateralization cues

Marcin Stachurski, Robert J. Summers, Brian Roberts*

Psychology, School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, UK

ARTICLE INFO

Article history:
Received 4 May 2017
Received in revised form
25 July 2017
Accepted 31 July 2017
Available online 2 August 2017

Keywords:
Verbal transformation effect
Auditory grouping
Concurrent speech segregation
Fundamental frequency
Interaural time difference
Within-ear interaction

ABSTRACT

Repeating a recorded word produces verbal transformations (VTs); perceptual regrouping of acoustic-phonetic elements may contribute to this effect. The influence of fundamental frequency (F0) and lateralization grouping cues was explored by presenting two concurrent sequences of the same word resynthesized on different F0s (100 and 178 Hz). In experiment 1, listeners monitored both sequences simultaneously, reporting for each any change in stimulus identity. Three lateralization conditions were used – diotic, ±680-μs interaural time difference, and dichotic. Results were similar for the first two conditions, but fewer forms and later initial transformations were reported in the dichotic condition. This suggests that large lateralization differences *per se* have little effect – rather, there are more possibilities for regrouping when each ear receives both sequences. In the dichotic condition, VTs reported for one sequence were also more independent of those reported for the other. Experiment 2 used diotic stimuli and explored the effect of the number of sequences presented and monitored. The most forms and earliest transformations were reported when two sequences were presented but only one was monitored, indicating that high task demands decreased reporting of VTs for concurrent sequences. Overall, these findings support the idea that perceptual regrouping contributes to the VT effect.

1. Introduction

It has long been known that repeating aloud a word to oneself over and over leads to the sound of that word losing its meaning (e.g., Titchener, 1915, pp. 26–27); this lapse in meaning is called verbal satiation. A closely related phenomenon is the verbal transformation effect (VTE), in which listeners report changes in verbal form when a recording of a spoken word is repeated many times (Warren, 1961a; for reviews, see Warren, 1996, 2008). The VTE involves a series of abrupt changes in the perception of the speech signal, some to new forms and others back to forms previously reported. Notably, these alternative forms often involve complex phonetic distortion of the stimulus. The VTE is not simply a laboratory curiosity; it can provide insights into how the auditory system processes ambiguous sensory information and switches between alternative interpretations of that information.

The changes in verbal form which characterize the VTE were originally interpreted mainly in terms of linguistic processes. Specifically, it has been argued that verbal satiation (adaptation) of a given form occurs once that form has been perceived for a time and a new perceived form emerges from among competing lexical candidates (or sometimes phonologically plausible non-words) as a result of criterion shift (Warren, 1968). These processes continue and the new form itself undergoes satiation, replacement, and recovery from adaptation. More generally, the VTE has been seen as related to changes in the perception of connected discourse that may occur when the initial linguistic interpretation is not confirmed by subsequent context (Warren, 1968), and hence as related to mechanisms normally used to resolve ambiguities and correct errors when listening to speech (Warren and Warren, 1970; Obusek and Warren, 1973; Kaminska et al., 2000; Basirat et al., 2012). In addition, the profound changes across the lifespan observed for the frequency and type of transformations reported by listeners are consistent with age-related changes in linguistic skills and experience (Warren, 1961b; Warren and Warren, 1966).

It has also long been recognized that the VTE shares some

Abbreviations: F0, fundamental frequency; ITD, interaural time difference; PSOLA, Pitch Synchronous Overlap and Add method; VTE, verbal transformation effect

* Corresponding author.
E-mail address: b.roberts@aston.ac.uk (B. Roberts).

http://dx.doi.org/10.1016/j.heares.2017.07.016

common features with the temporal characteristics of the shifts in perceptual organization associated with reversible and multistable visual figures (e.g., Warren and Gregory, 1958; Ditzinger et al., 1997); indeed, recent research using neuroimaging has provided evidence that common functional brain networks underlie perceptual switching in auditory streaming and in verbal transformations (Kashino and Kondo, 2012). However, it is only since the millennium that the relationship between the VTE and cues for auditory stream segregation (Bregman and Campbell, 1971; Bregman, 1990) has been explored in any detail. Pitt and Shoaf (2002) showed that the verbal transformations experienced by listeners are related to the acoustic cues that help bind together the rapidly changing and diverse sounds of speech (see, e.g., Darwin, 2008). In particular, acoustic-phonetic elements that are periodic and have a low-frequency centroid, such as nasals, cohere better with neighbouring voiced vowels than do acoustic-phonetic elements that are aperiodic and have a high-frequency centroid, such as unvoiced fricatives, affricates, and plosives. Hence, extended repetition tends to lead to segregation of unvoiced consonants from the core vocalic parts of the stimulus into one or more streams, with the reported verbal form corresponding to the foreground percept and the unreported segments corresponding to the background. Therefore, the VTE is influenced not only by linguistic processes, but also by the cleaving off and regrouping of acoustic-phonetic elements in a speech stimulus. Since the establishment of this relationship, the VTE has been used as a means of investigating the role of formant transitions and the continuity of the pitch contour in holding together the speech stream (Stachurski, 2012; Stachurski et al., 2015) and of the role of lexical knowledge in the formation of speech streams (Billig et al., 2013).

All of the VTE studies considered so far involved presenting listeners with only one repeating stimulus sequence at a time. Warren and Ackroff (1976) adapted the established methods for studying the VTE to examine the effects of presenting two identical sequences at once (see also Warren, 1996). Fig. 1 illustrates the dichotic stimulus configuration used; the two sequences were distinguished by ear of presentation and played half a cycle out of phase to prevent binaural fusion. One aim of that study was to establish whether or not the same transformations would be heard at the same time on the left and the right; another was to explore the effect of the task demands involved when listeners monitor both sequences at once. It was assumed that simultaneous and identical changes would indicate that a single set of linguistic units was involved in processing both sequences, whereas independent changes would indicate two (or more) sets of functionally separate units. Warren and Ackroff reported that changes occurred at different times at the two ears and that all listeners had periods of time during which they perceived two different forms. For example, a repeating sequence of the word "tress" might be heard at a particular time as "dress" in one ear but as "commence" in the other. That listeners heard independent changes at the two ears was taken to indicate the involvement of more than one set of linguistic units in processing the two sequences, suggesting that everyday listening under cocktail-party conditions (Cherry, 1953) typically involves the processing of speech arising from spatially distinct sources by independent lexical analysers. Note, however, that the method used did not include an accurate measure of time – the results for each trial consisted only of an ordinal list of transcribed responses flagged with the ear to which the listener was responding. Hence, the *degree* of independence in the responses to the two sequences was not quantified.

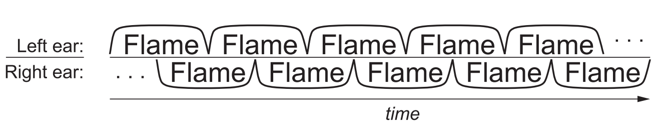

Fig. 1. Schematic illustrating the experimental setup used by Warren and Ackroff (1976) for the dichotic presentation of sequences of repeating stimulus words one half-cycle out of phase, using the example word "flame".

Warren and Ackroff (1976) also briefly reported a preliminary study in which five experienced listeners heard three concurrent sequences of the same stimulus; these sequences were each offset by one-third of a cycle – one to the left ear, one to both ears (centre), and one to the right ear. Monitoring all three sequences at once was challenging even for these experienced listeners, but all of them reported independent changes at the different spatial positions. To our knowledge, only one full-scale study has followed up on these observations (Zuck, 1992). That study used a similar configuration of three sequences but listeners were asked to monitor only one or other of them, and so the results did not provide any further insight into the independence of the transformations heard across the three sequences.

Our current understanding of the relationship between the VTE and auditory stream segregation, which is based on studies using single sequences (e.g., Pitt and Shoaf, 2002; Stachurski et al., 2015), suggests that stimulus configurations that increase the possibilities for perceptual regrouping of acoustic-phonetic elements should facilitate the VTE. Furthermore, it also suggests an alternative or additional explanation for the independent verbal transformations for two concurrent sequences of identical stimuli observed by Warren and Ackroff (1976). Specifically, different transformations at the two ears might be a consequence, at least in part, of independent streaming processes at the two ears that lead to independent patterns of segregation and regrouping for the acoustic-phonetic constituents of the stimulus. If this is the case, then using a stimulus configuration that lowers the likelihood of independent changes in the perceptual organization of the two sequences, relative to dichotic presentation, should decrease the independence of the verbal transformations reported for those sequences. Two experiments are reported here.[1] The first tested the hypothesis that allowing two concurrent sequences to interact in the auditory periphery would facilitate the VTE by increasing opportunities for perceptual reorganization but would also decrease the independence of the responses to the two sequences. The second experiment further explored the effects on the VTE of peripheral interaction between sequences and also extended Warren and Ackroff's research on the impact of task demands on listeners' responses to concurrent sequences.

2. Experiment 1

This experiment compared the patterns of verbal transformations for two concurrent sequences under dichotic presentation, in which the two sequences were isolated from one another at the auditory periphery, with those for conditions in which both sequences were present in both ears, in which the acoustic-phonetic constituents of the two sequences were able to interact in the same ear. The two voices were always distinguished using differences in fundamental frequency (F0), either with or without an additional lateralization cue based on ear of presentation or on interaural time difference (ITD) cues; ITD cues help listeners to track the speech of a particular talker across time (Darwin and Hukin, 1999). Differences in F0 provide a salient concurrent segregation cue known to be important in separating speech mixtures (Brokx and Nooteboom, 1982; Bird and Darwin, 1998;

[1] The experiments reported here correspond to reanalysed versions of experiments 1 and 2 in the doctoral thesis of Marcin Stachurski.