3235 Kifer Rd. • Suite 100
Santa Clara, CA 95051

www.dspconcepts.com

DESIGNING OPTIMIZED MICROPHONE BEAMFORMERS

Our previous paper, "Fundamentals of Voice UI," explained the algorithms and processes required for a voice UI system. In this paper, we demonstrate how the different microphone types and array configurations affect performance of voice UI systems, and make specific recommendations engineers and product design teams can use to get the best performance from their voice UI products.

As we discussed in the previous paper, beamforming is the processing of signals from multiple omnidirectional microphones to focus in on the sound coming from the direction of the most prominent source (i.e., the user's voice) and disregard sounds coming from other directions. A direction-of-arrival (DOA) algorithm first determines the desired direction of beam focus, then the beamformer algorithm passes the sound from the nearest microphone while manipulating the phase of the signals from the other microphones so sounds coming from outside the beam are reduced in level.

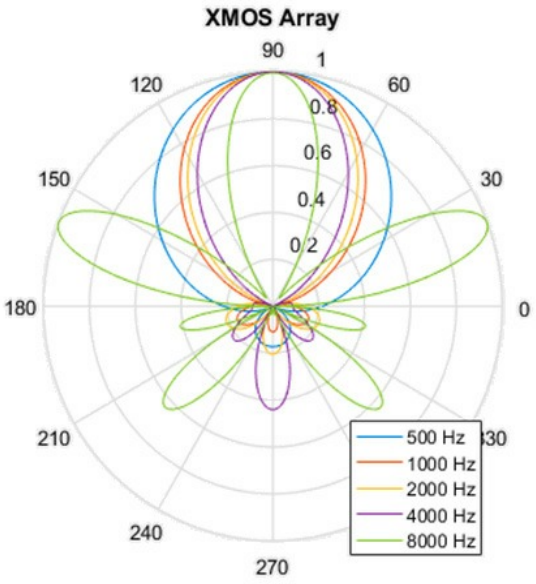

***Figure 1:* Pickup pattern of seven-microphone array similar to the one used in the Amazon Echo, tested from an angle of 90°. There is one microphone in the center and six microphones evenly spaced around the center mic at a radius of 30mm.**

The traditional way of evaluating microphone beamformers is to look at their beam patterns. These graphs indicate how the array works as a spatial filter, and its ability to enhance desired sounds and remove undesired sounds. The figures below illustrate the performance of a seven-microphone array from 90 degrees off-axis. The design goal was to have a beamwidth of 45 degrees. The first plot indicates the array's beam pattern for different frequencies. Note that the array becomes more directional as frequencies increase from 500 Hz to 4 kHz. At 8 kHz, unwanted beams appear at roughly 20 degrees and 160 degrees. This is called spatial aliasing, and it occurs when the wavelength of the sound is smaller than the spacing between microphones.

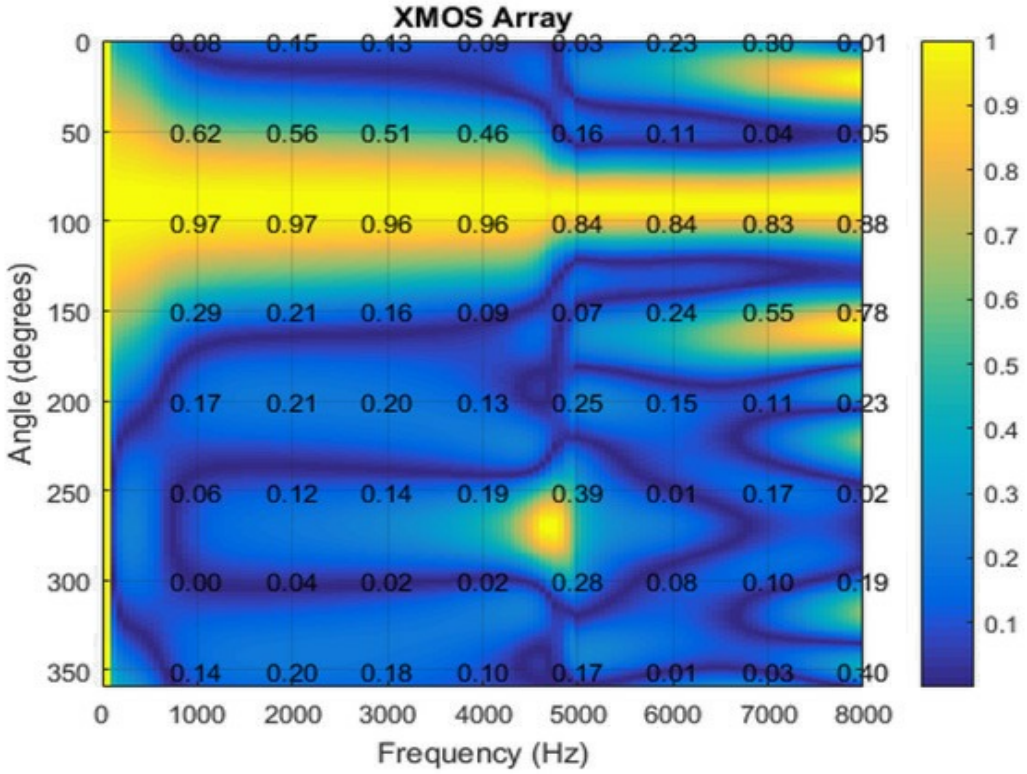

***Figure 2:** Beamforming performance of a microphone array from a look angle of 90° off-axis. Yellow sections indicate the strongest reception of sound, while dark blue areas indicated the strongest rejection of sound.*

The graph above shows another way of visualizing the same information. Instead of showing a polar pattern, the spatial information is presented on a two-dimensional grid. The Y axis corresponds to the look angle (the angle between the user and the front of the device), and the X axis corresponds to frequency. The color indicates how much attenuation the array provides in each direction. Note that the array is omnidirectional at low frequencies and narrows at high frequencies. The unwanted high frequency beams also appear at about 5 kHz.

The question to answer is, how does the performance of this mic array translate into accuracy of voice recognition?

Microphone arrays are designed using sophisticated mathematical algorithms. Most of these algorithms focus on optimizing the beam pattern of the array, with the hope that this will translate into actual improvements in performance.

Our experience at DSP Concepts has been that the beam pattern is just one element of voice UI system performance that must be considered. We have found that the ratio of the user's voice to the background noise is the ultimate determiner of the performance of a voice UI system. This is like a signal-to-noise ratio; the signal is the level of speech and the noise corresponds to other interfering sounds in the room. The performance of wake word algorithms, beamformers, and AECs all correlate to the signal-to-noise ratio.

"We have found that the ratio of the user's voice to the background noise is the key determiner of the performance of a voice UI system."

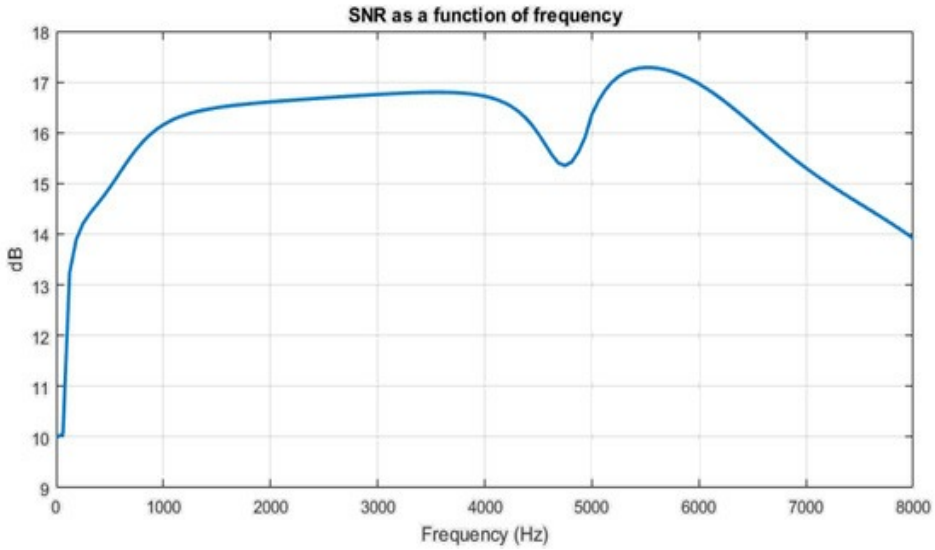

***Figure 3:** Signal-to-noise performance of a typical multiple-microphone array using beamforming*

So instead of looking at beam patterns, at DSP Concepts we focus on the SNR of an array. For example, if we assume a speech level of 60 dB SPL, a background noise level of 50 dB, and a microphone SNR of 64 dB, then the SNR at the output of the array is shown in **Figure 3** above.

This setup has an intrinsic SNR of 10 dB (the speech is 10 dB louder than the background noise) and this is the performance you would achieve with a single microphone. Any SNR improvement above 10 dB is due to the microphone array, and the graph shows roughly 6 dB improvement above 1 kHz. The SNR then starts dropping above 4 kHz and this corresponds to the spatial aliasing that we discussed earlier.

Evaluating based on SNR provides intuition into how much benefit the array will bring. A 6 dB improvement allows you to "stand 6 dB further away" – or twice as far – as would be possible with a single microphone.

As we noted in our previous paper, an improvement of just a few dB in signal-to-noise ratio may seem insignificant to many audio professionals, who are accustomed to SNRs that are typically far better than the application demands. For example, improving an amplifier's SNR from 105 dB to 109 dB will not result in a subjectively appreciable performance improvement.

However, SNR is much more important in voice UI applications, where the user's voice may be at the same level as the surrounding noise or the music playback coming from the speaker that houses the voice UI system. Thus, microphone configurations and processing algorithms that can elevate a user's voice just a few additional dB above the environmental noise can produce a large improvement in voice-recognition accuracy.

Parameters Affecting Array Performance

This section presents the SNR performance of various microphone array designs. Parameters evaluated here are:

1) Number of microphones
2) Microphone geometry and spacing
3) Background noise level
4) Microphone noise floor (its SNR)

We studied an array in which the microphones were arranged on a circle with a 71mm diameter. Rated SNR of these mics was 64 dB. Testing was done in an environment with diffuse-field noise at 50 dB SPL, with a speech signal at average 60 dB SPL. Beamwidth was 45 degrees, and look angle was 0 degrees except where otherwise specified. Signal processing was performed using DSP Solutions' Audio Weaver Voice UI algorithm package.

Test #1: Number of Microphones

In the chart below, the six-microphone array shows a clear advantage in signal-to-noise ratio at all frequencies from 1000 to 5500 Hz. As the number of mics is reduced, overall SNR suffers, although reducing the number of microphones can actually improve SNR within certain frequency bands.

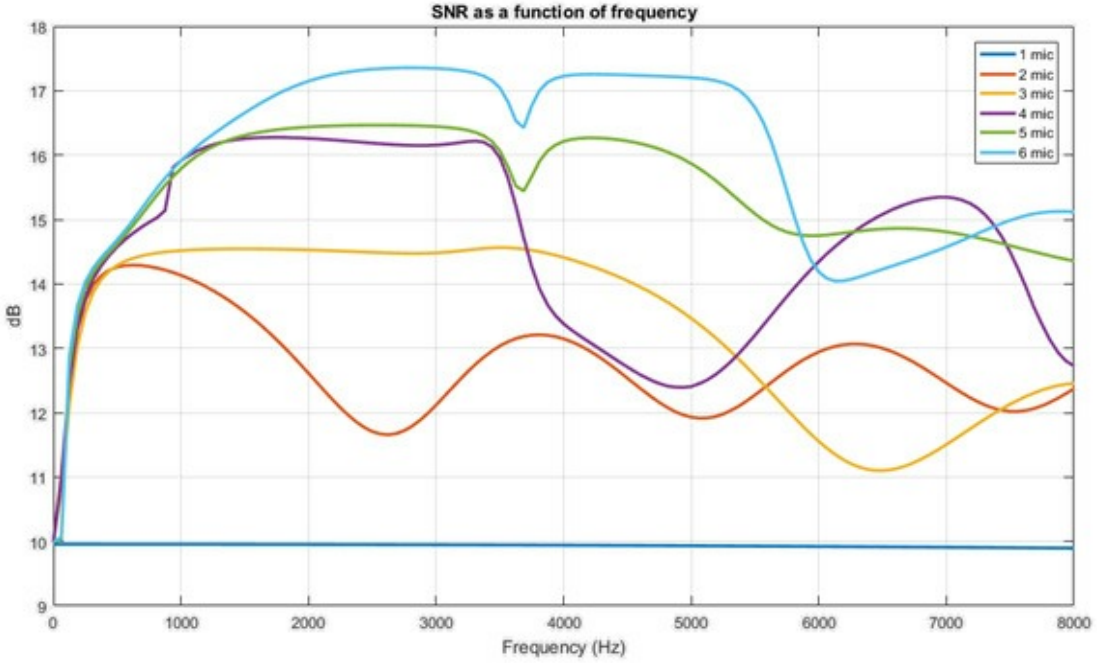

Figure 4: ***Signal-to-noise ratio of single microphone compared with arrays of two to six microphones. The higher the line on the chart, the better the SNR.***

The polar plots below and on the following page show the pickup patterns of circular arrays using one to six microphones, measured at frequencies of 500, 1000, 2000 and 4000 Hz (thus covering most of the range of human speech). Ideally, the pickup pattern should show a tight beam pointed directly to the right, at the look angle of 0°, with little variation at different frequencies.

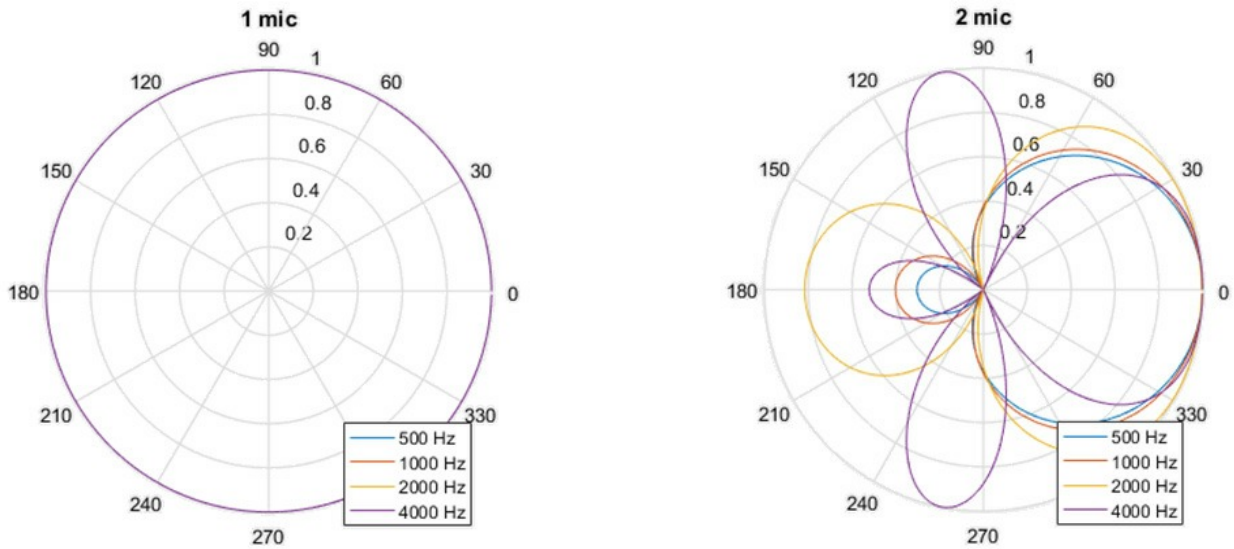

Figure 5: ***Pickup pattern at four different frequencies for single omnidirectional microphone (left) and two-mic array (right)***

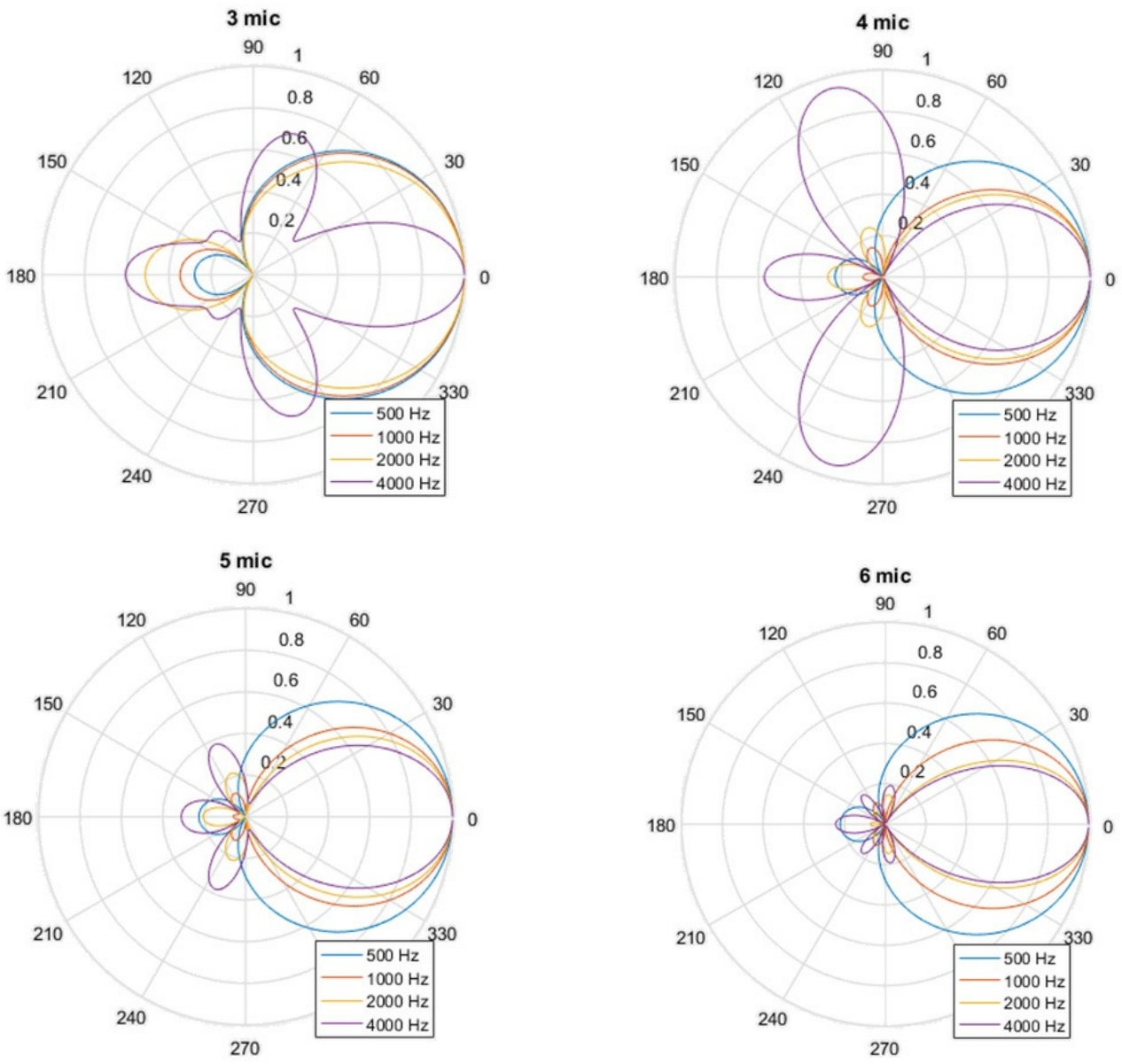

***Figure 6:** Pickup pattern at four different frequencies for (clockwise from top left) three-, four-, five, and six-microphone arrays*

As can be seen in the polar plots, increasing the number of microphones generally allows for a tighter, more focused pickup beam, but in certain cases adding microphones does not improve performance at all frequencies. For example, three microphones clearly produce a better result at all frequencies than two microphones, however, increasing the microphone count to four improves performance from 500 to 2000 Hz but degrades it at 4000 Hz. The two-, three- and four-microphone arrays produce significant off-axis lobing at 4000 Hz; this reduces system SNR and increases the chance of an inaccurate DOA determination, which could make the beamformer aim in the wrong direction.

The two-microphone array, in particular, does a relatively poor job of rejecting sounds from 180°. (The three- and four-mic arrays also exhibit this flaw, but only to a significant amount at 4000 Hz.) This error can be especially problematic if the unit is placed near a wall or other large sound-reflecting object, where the reflection might cause the voice UI system to think the user's voice is coming from the wall instead of from the user.

The arrays of five and six microphones produce better results, with tightly focused beams on the 0° axis, negligible off-axis lobing, and excellent rejection of sounds from 180°.

Test #2: Microphone SNR

Because system SNR is critical to accurate voice recognition, it's tempting to assume that using microphones with higher SNR would improve voice UI performance. To test this assumption, total system SNR was tested with microphones rated at 64 and 70 dB SNR, each type arranged in arrays comprising one to six microphones. The main advantage of a high-SNR mic would occur at low frequencies, because the improved SNR would permit more aggressive processing of low frequencies, which is where most environmental noise in homes and autos typically occurs.

The following two graphs show how microphone signal-to-noise ratio affected the performance of the different microphone arrays, with system SNR shown relative to frequency. The higher the trace is on the chart, the better the SNR and the better the performance of the voice UI system should be. Solid lines show the results with the 64 dB SNR microphone; dotted lines show the result with the 70 dB SNR microphone. The tests were performed twice: once with 50 dB SPL ambient noise, once with 35 dB ambient noise.

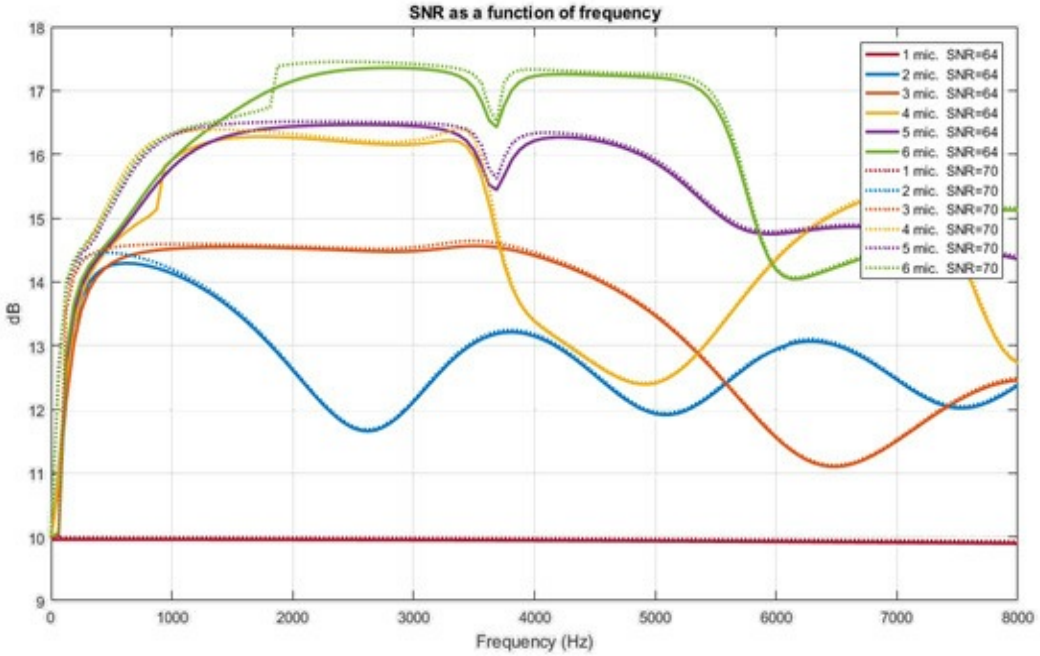

***Figure 7:* Comparing the system SNR of one- to six-microphone arrays implemented with standard (64 dB SNR) mics and low-noise (70 dB SNR) mics, measured in a test environment with 50 dB SPL ambient noise**

The graph above shows the results with a 50 dB SPL ambient noise field, which is what would be encountered in a typical residential living room with common levels of noise from appliances, pets, light conversation in other rooms, etc. In this case, the improvement gained by using microphones with better SNR is in most cases barely measurable and would not noticeably improve voice UI performance.

The graph on the next page shows the same test conducted in a background noise level of 35 dB, which corresponds to a very quiet home environment. Under these conditions, using microphones with better rated SNR has a larger impact, in many cases increasing system SNR by about 1 dB. However, note that reducing ambient noise has a much larger impact on system SNR, typically improving it by about 14 dB. Thus, the benefits of a 1 dB improvement in mic SNR on overall system performance would be insignificant in this case.

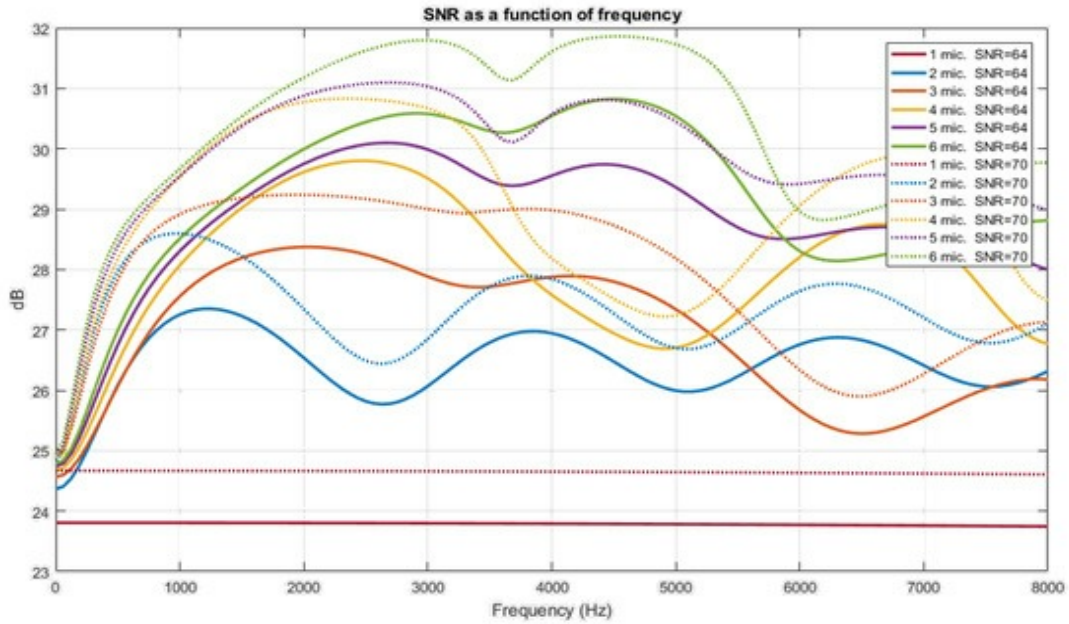

Figure 8: ***Comparing the system SNR of one- to six-microphone arrays implemented with standard (64 dB SNR) mics and low-noise (70 dB SNR) mics, measured in a test environment with 35 dB SPL ambient noise***

Test #3: Microphone Gain Matching

Like other mechanical devices, microphones exhibit unit-to-unit inconsistency. The gain of two samples of the same microphone can vary substantially; a tolerance of ±3 dB (for a maximum difference of 6 dB in gain between two samples) is common. In arrays of multiple microphones, these inconsistencies might negatively affect system SNR and the overall performance of the voice UI system. Microphones with tighter gain tolerance, or with factory calibration measurements for each mic, are sometimes available, but they are typically more costly.

To evaluate the effects of microphone gain mismatch on system SNR, models of theoretical arrays of one to six perfectly matched microphones were tested. Gain mismatches of ±1, ±2 and ±3 dB were then introduced into the model, and the tests repeated.

The three graphs shown in **Figure 9** on the next page show how microphone gain tolerance affected the performance of the different arrays, with system SNR shown relative to frequency. The higher the trace is on the chart, the better the SNR and the better the performance of the voice UI system should be. Solid lines show the results with perfect gain matching; dotted lines show the result with the gain mismatched at ±1 (first chart), ±2 (second chart) or ±3 dB (third chart).

These charts show that gain mismatches in arrayed microphones can have a large negative impact on system SNR, often comparable to the impact that reducing the number of microphones might have. The effect is particularly noticeable in the bottom chart, with the ±3 dB mismatch that is typical of the microphones used in voice UI systems.

"Compared with using low-noise microphones, reducing ambient noise has a much larger impact on system SNR, typically improving it by about 14 dB."

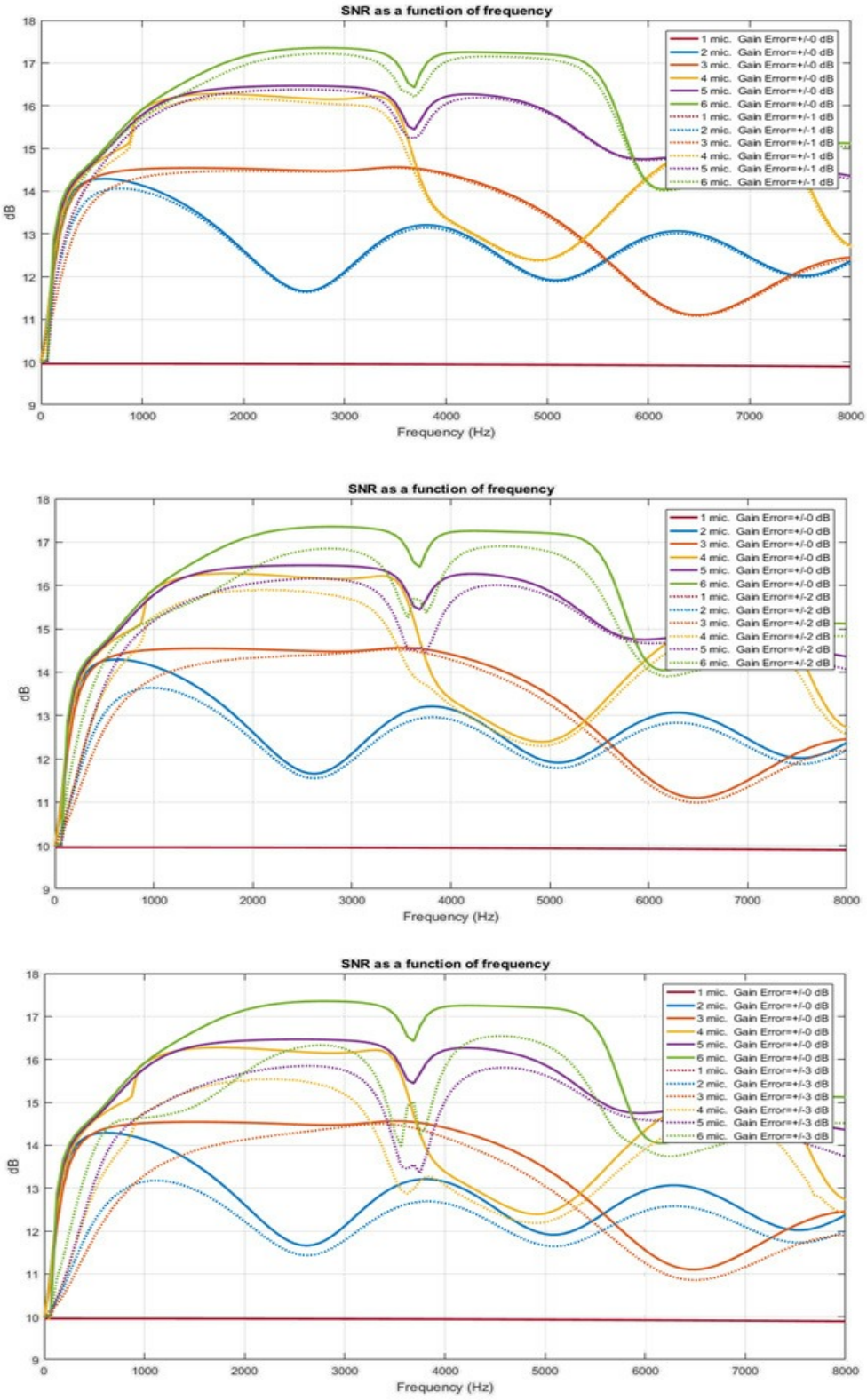

Figure 9: ***Comparing the system SNR of one- to six-microphone arrays implemented with mics matched to a tolerance of ±1 dB (top), ±2 dB (middle) and ±3 dB (bottom)***

These tests were performed on a theoretical array without an enclosure. Once the mics are mounted in an enclosure, gain and frequency response will change depending on how and where the mics are mounted and the consistency of the acoustic seals around the mics. For this reason, using mics of better consistency, or supplied with factory calibration data, may not produce an optimal result because the acoustical effects of the enclosure and mounting may introduce performance inconsistencies even with the most tightly matched microphones.

The best solution in this case is for the microphone gain to be measured with the mics installed, and the gain for each mic adjusted in software. Ideally, each unit would be individually measured and calibrated in the factory after the product is assembled, so that the software can compensate for any inherent gain mismatch in the mics as well as for mismatches caused by the acoustical effects of the enclosure.

Test #4: Microphone spacing

Increasing mic spacing in an array might be expected to create greater differences in level among all the mics because the source-to-mic distances will be greater. It will also alter the relative phase among the mics. To find out how spacing affects SNR, arrays using two to six mics were tested, with mics placed on circles ranging from 5 to 71mm in diameter. A three-mic array was tested with the mics placed on circles measuring 40, 80, 160 and 320mm.

The following eight graphs show how mic spacing affected array performance, with system SNR shown relative to frequency. The higher the trace is on the chart, the better the system SNR and the better the performance of the voice UI should be. Results for the two- and three-microphone arrays are shown first from an end-fire position (with the source directly in line with one of the mics) and a broadside position (with source equidistant from two microphones).

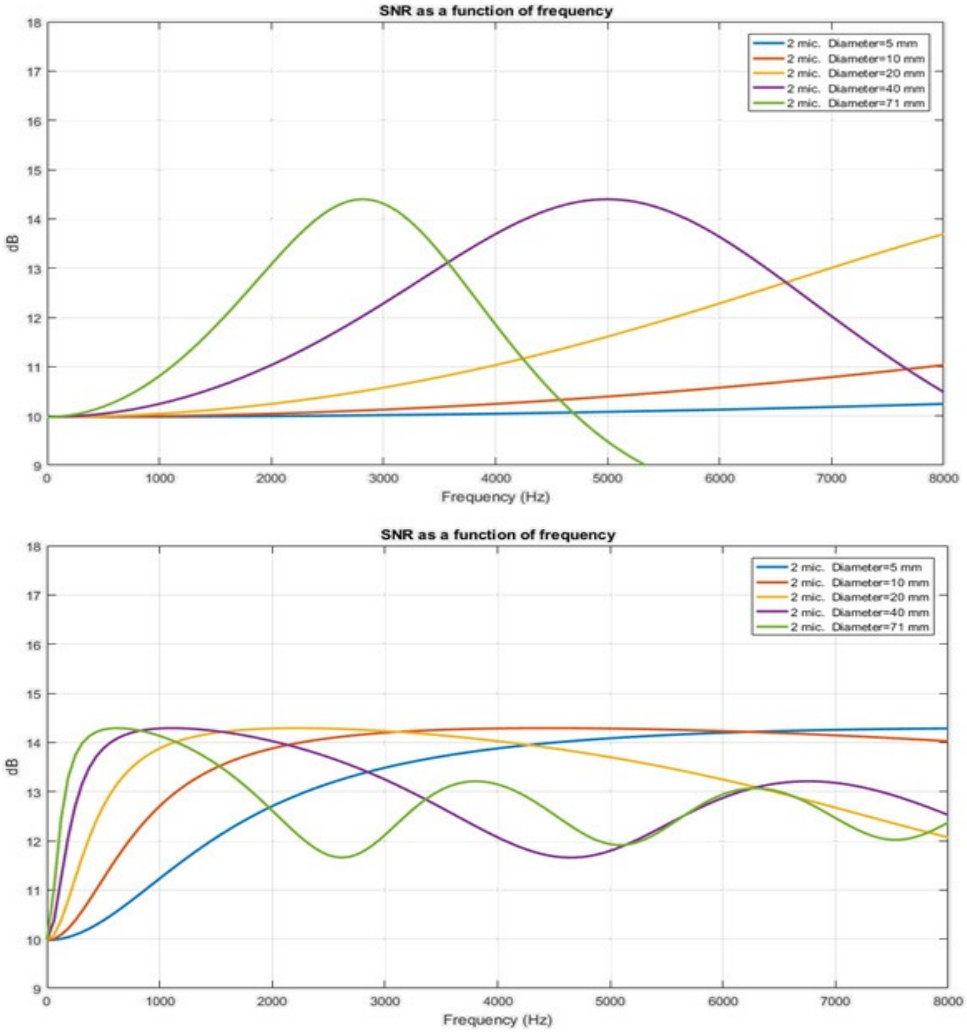

Figure 10: ***Effect of different microphone spacings on a two-mic array, measured from end-fire (top) and broadside (bottom) positions***

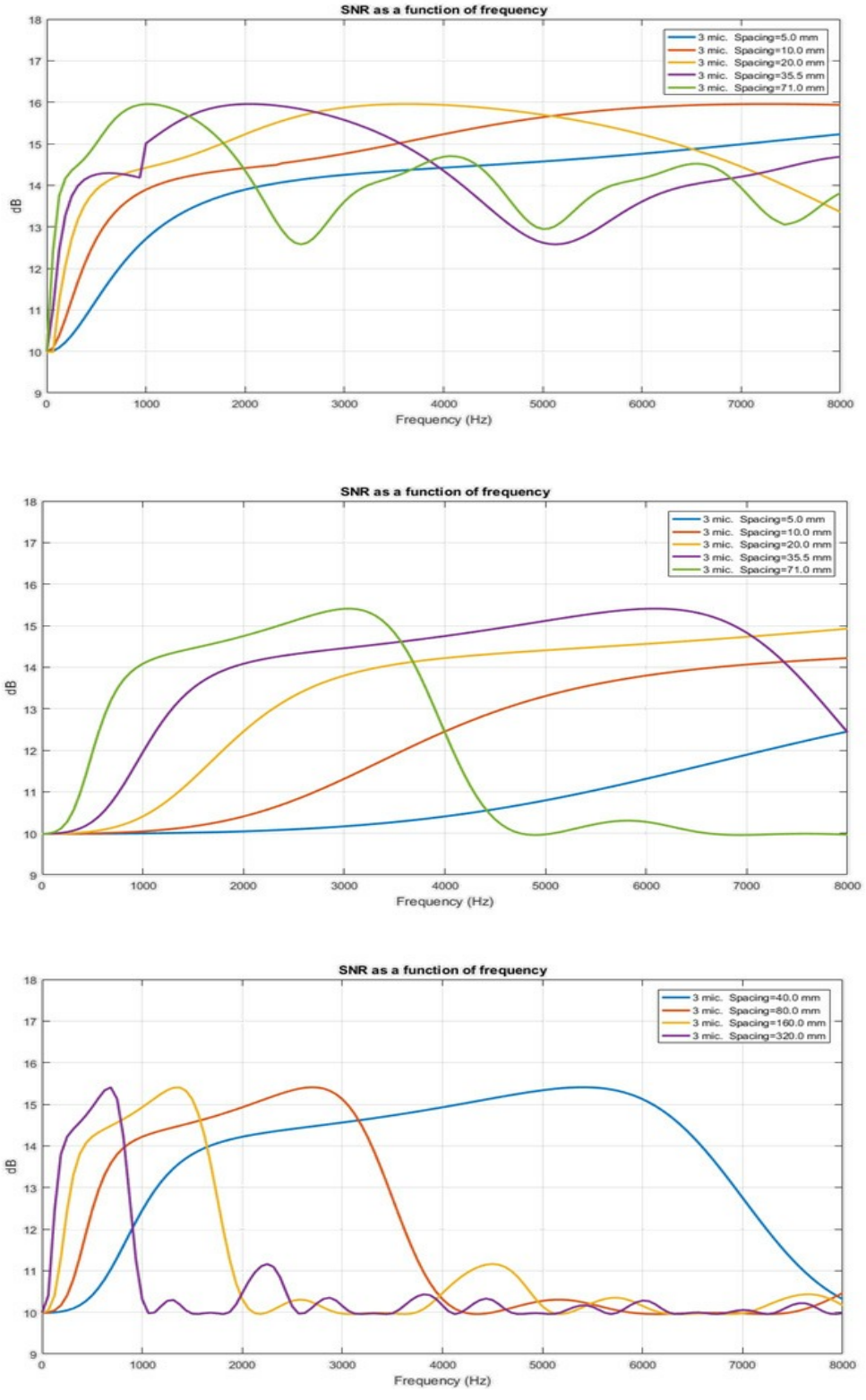

Figure 11: *Effect of different microphone spacings on a three-mic array, measured from end-fire (top) and broadside (middle) positions, and from broadside position on a three-mic array with wider mic spacings up to 320mm (bottom)*

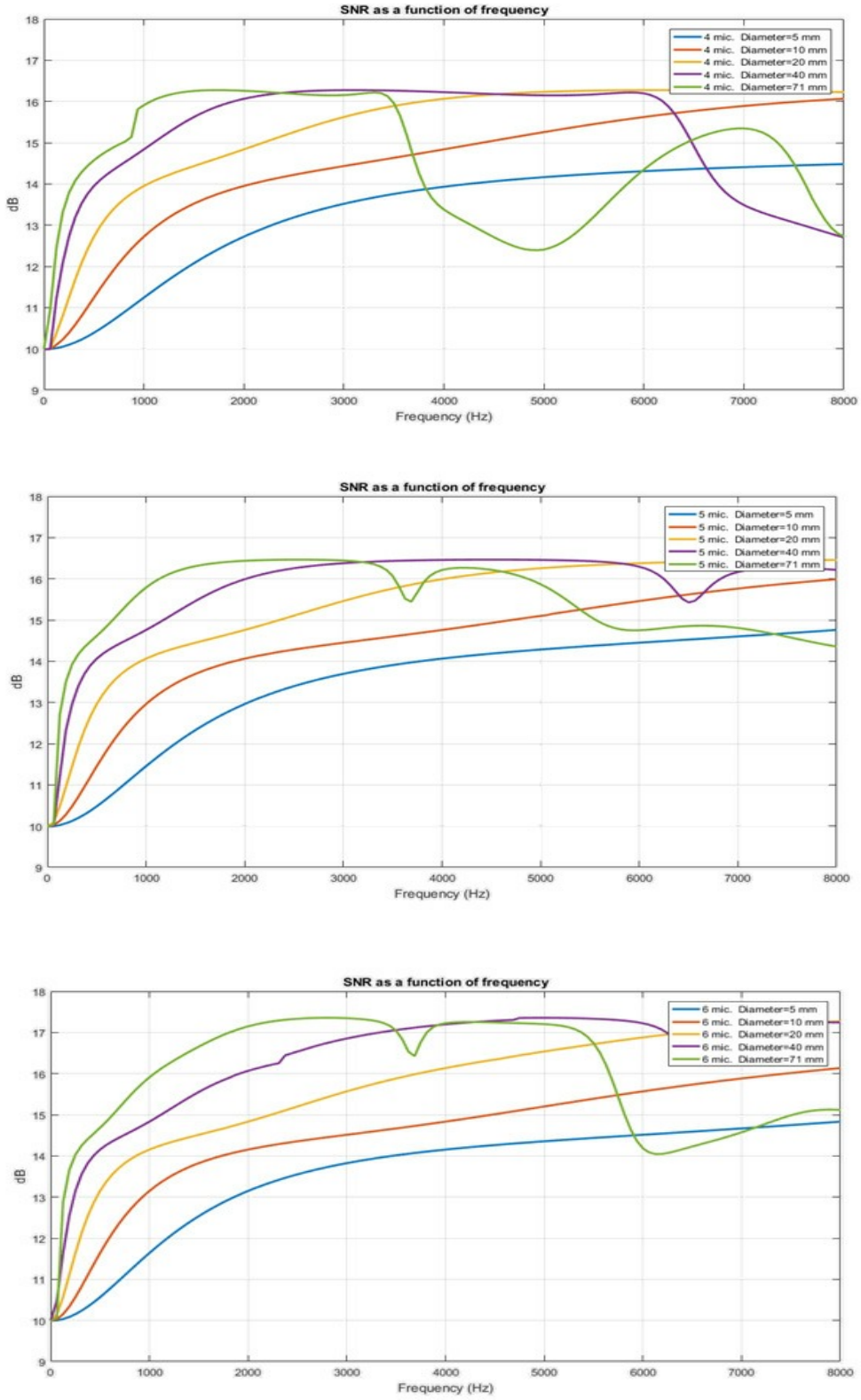

Figure 12: ***Effect of different microphone spacings on four-mic (top), five-mic (middle) and six-mic (bottom) arrays***

To summarize the results from the microphone spacing tests, within the human vocal range:

Two-mic array: Best results at 71mm spacing from end-fire position, 40mm from broadside.

Three-mic array: Best results at 35.5mm from end-fire position and 71mm from broadside position. With wider spacings, increasing spacing to 80mm further improves performance, but greater increases narrow the range of optimum SNR to the point where SNR in some of the vocal range would decrease.

Four-mic array: Best results at 71mm.

Five-mic array: Best results at 71mm.

Six-mic array: Best results at 71mm, although performance at 40mm is not as close to the result at 71mm as it is with the four- and five-mic arrays.

Based on these results, placing microphones on a circle measuring 71mm in diameter is generally the best choice with arrays of three to six microphones, if there is sufficient physical space on the voice UI device. With a two-microphone array, results vary considerably depending on whether the source is in line with the microphones or broadside to the mic array. This result suggests that if a two-microphone array is used (either to cut costs or accommodate specific device form factors), an end-fire configuration should be used if at all possible. The broadside configuration yields very little SNR improvement.

Performance of Existing Products

While it is important for a product team to understand how design decisions affect performance of a voice UI system, it is also useful to have benchmarks against which a product can be compared. For this reason, the performance of the mic arrays and algorithms of three existing products were tested. These products are the Amazon Echo and Google Home smart speakers. Also included are measurements of a system using DSP Concepts Voice UI algorithm.

Amazon Echo

The Echo smart speaker has a top-mounted array of seven microphones—one centered with six arranged in an 82mm circle around it, along the top outer edge of the chassis.

The following graphs show how well the Echo is able to focus on sound from a certain direction. These measurements were taken at a look angle of 90°, perpendicular to the Amazon logo on the front of the unit. The polar pattern below left shows a fairly tight beamwidth averaging about ±30°, with very little signal picked up from the rear and only slight side lobing at 8000 Hz. The chart at right shows how the beamwidth varies with frequency. The chart on the next page shows that the system SNR of the Echo's mic array and processing is relatively high, and relatively consistent at most frequencies.

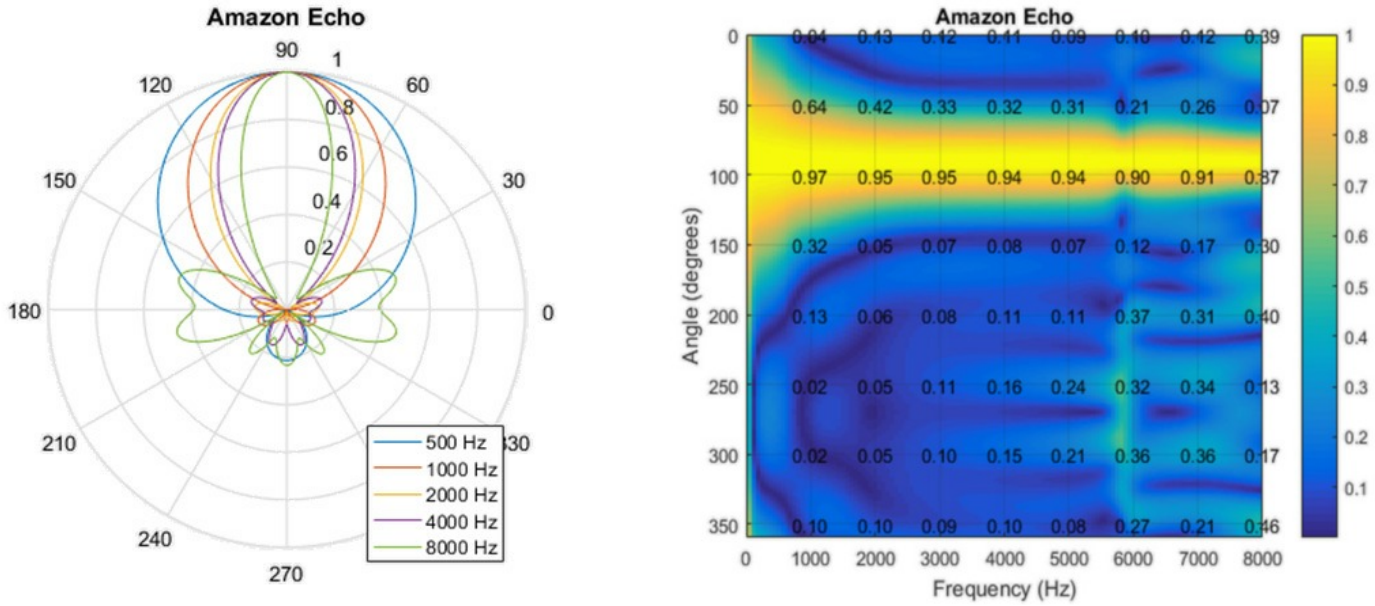

***Figure 13:** Polar pickup pattern (left) and beamforming performance (right) of original Amazon Echo.*

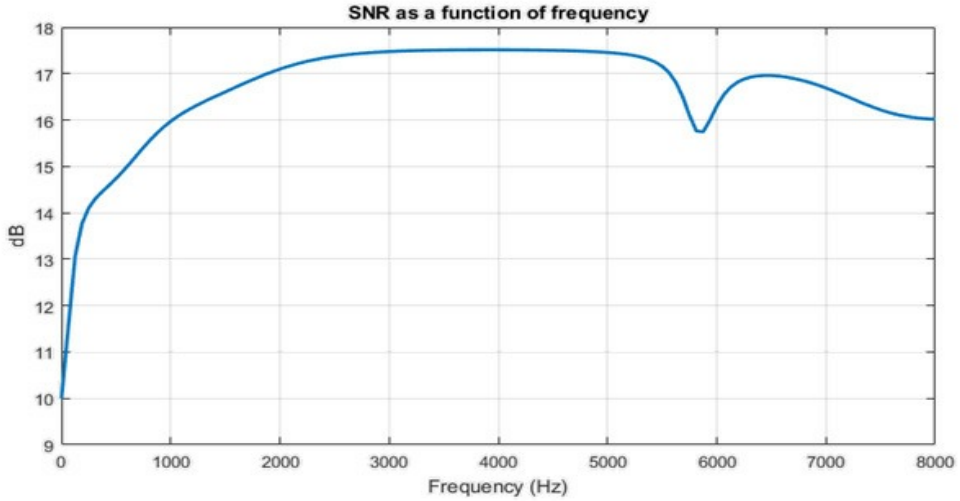

Figure 14: ***System SNR vs. frequency of Amazon Echo***

Google Home

The Home smart speaker has an array of two microphones, one at each side of its slanted top, spaced 71mm apart. The unit is designed so that a look angle of 0° addresses the two microphones broadside. The following graphs show how well the Home is able to focus on sound from look angles of 0° (broadside) and 90° (end-fire).

The chart below left shows the polar pattern of the Home's mic array from the broadside (0°) position, which shows strong lobing at 4000 and 8000 Hz, a figure-8 pickup pattern at 2000 Hz and a nearly omnidirectional pickup pattern at lower frequencies. The chart below right shows that the directionality of the array varies greatly with frequency.

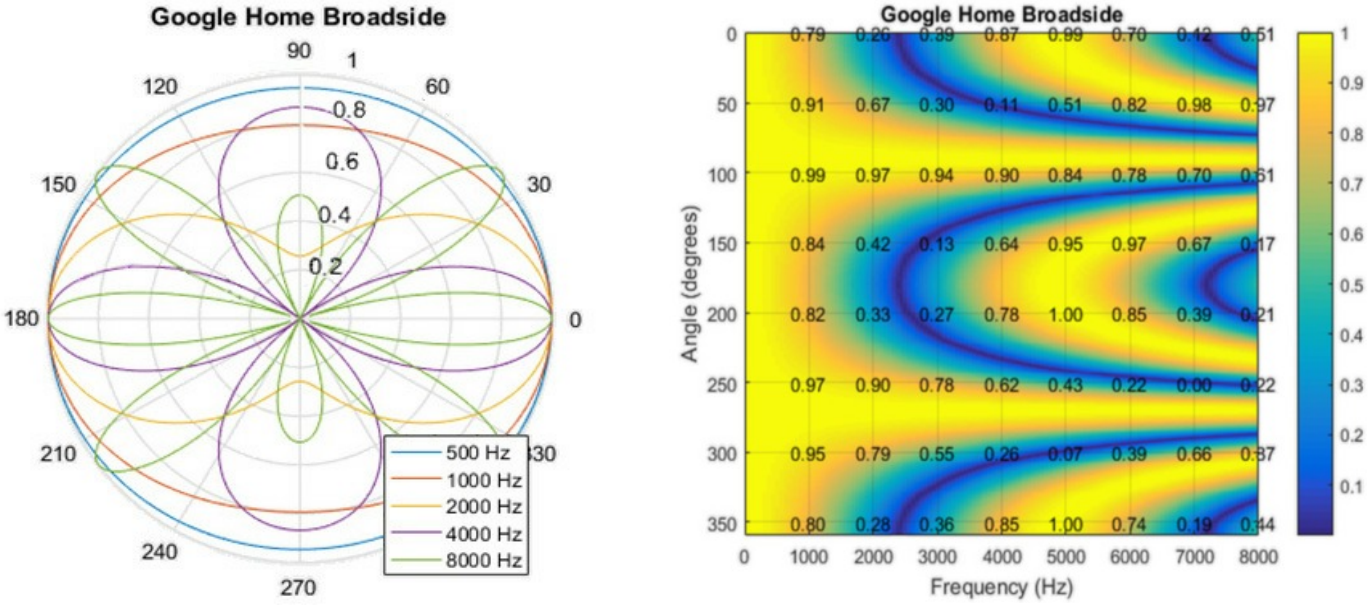

Figure 15: ***Polar pickup pattern (left) and beamforming performance (right) of Google Home smartspeaker tested from broadside (0°) position***

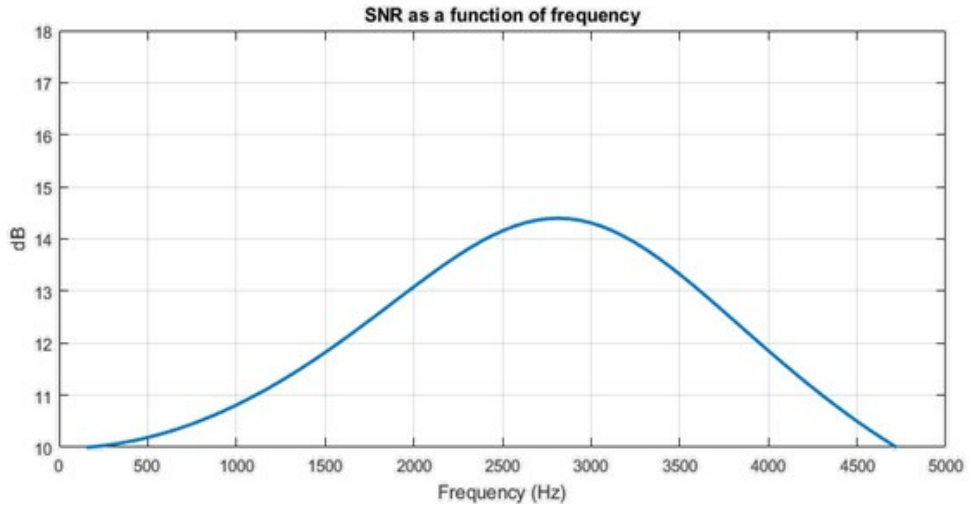

Figure 16: ***System SNR vs. frequency of Google Home measured from broadside position***

The chart above shows the system SNR of the Home's mic array from the broadside (0°) position. The SNR is optimal only within a narrow range, and is less than that shown in most of the other measurements performed for this paper.

The chart below left shows the polar pattern of the Home's microphone array from the end fire (90°) position, which shows considerable lobing at 4000 and 8000 Hz, with strong pickup from 270°. The chart below right shows that the pickup is strongest from 90° but also relatively strong from other directions as well, indicating that the array's directionality is not well-focused.

The chart below right shows the system SNR of the Home's mic array from the end fire (90°) position, which is relatively poor and somewhat inconsistent as frequency changes.

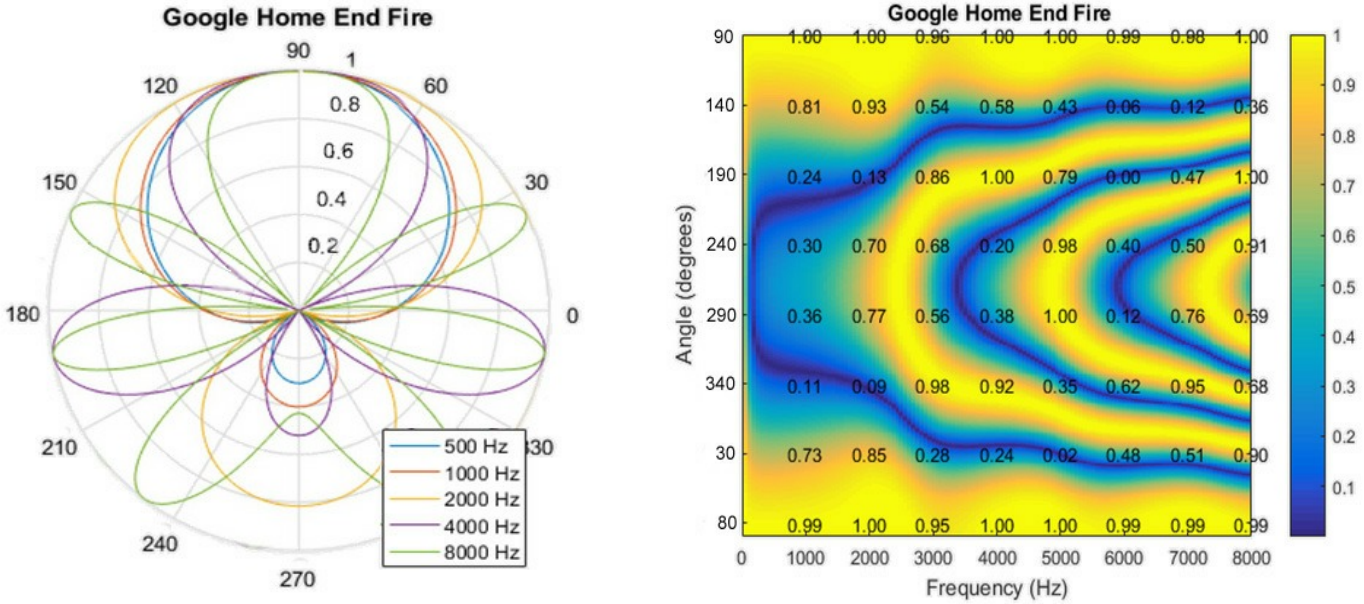

Figure 17: ***Polar pickup pattern (left) and beamforming performance (right) of Google Home smartspeaker tested from end-fire (90°) position***

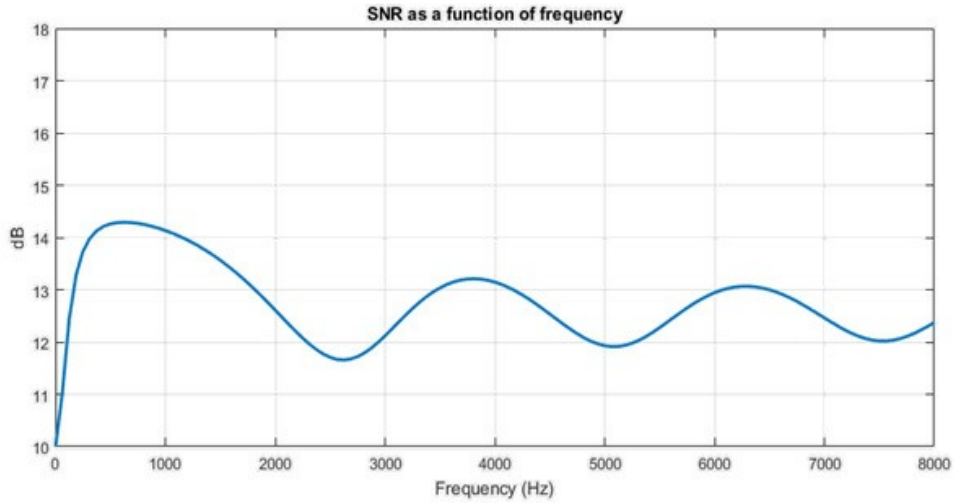

Figure 18: ***System SNR vs. frequency of Google Home measured from end-fire position***

Conclusions

The most important conclusion to draw from this paper is that there is much product development teams can do to optimize the performance of voice recognition systems. By giving the voice recognition AI system the cleanest possible voice signal to work with, engineers can assure the most accurate voice recognition and the greatest customer satisfaction.

Here are a few key principles product development teams should keep in mind when designing products that include voice command features:

One: Accuracy and reliability of voice recognition are the result of many factors, including form factor of the device, the components chosen, and the algorithms used. Excellent performance in one of these factors does not guarantee reliable performance of the entire system.

Two: Generally, the more microphones a product employs for voice pickup, the better. Three to six mics are optimal, although a five-mic array comes close to the maximum possible performance. Four microphones delivers a good balance of cost and performance.

Three: Microphone spacing of 40 to 80mm is typically best, with 71mm a good all-around choice that can be easily implemented in a wide range of products and applications.

Four: Matching the gain of the microphones in an array can substantially improve signal-to-noise ratio. Rather than using microphones with matched or calibrated gain, it is best to match the gain of the microphones at the factory through software, after they are installed in the voice-controlled device.